EVALUATION OF CARBON METABOLISM IN ENGINEERED NICOTIANA TABACUM FOR TERPENE BIOPRODUCT UNDER ENVIRONMENTAL VARIANCE

A Dissertation

by

CONNOR RILEY GORMAN

Submitted to the office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Joshua S. Yuan
Committee Members,	Herman B. Scholthof
	Michael V. Kolomiets
	Bhimanagouda Patil
Head of Department,	Leland S. Pierson III

August 2019

Major Subject: Plant Pathology

ABSTRACT

Terpenes and terpenoids have a vast array of applications for human use as fuels, specialty chemicals, nutraceuticals, pharmaceuticals, and food and fragrance additives. In plants, terpene production is limited and highly regulated, thereby restricting the capacity to use them as a source of terpene bioproduction. Many efforts using synthetic biology have been made to produce terpenes and terpenoids by metabolic engineering. This research focuses first on the biological limitations of current strategies to increase terpene production in higher order plants using carbon rerouting to an engineered terpene carbon sink, and the role the environment has on regulating bioproduction efficacy. This carbon rerouting seeks to utilize carbon from the byproduct of photorespiration, P-glycolate, for terpene synthesis by rearranging the two carbon molecule to a three carbon molecule of pyruvate which is used to produce DXP from condensation reaction with G3P. This reliance on light reactions means available carbon to synthesize terpenes became highly environmentally dependent. Then, we suggest an alternative carbon rechanneling route that increases carbon flux efficiency to terpene synthesis while reducing environmental impact on yields. The C2 redirection pathway metabolome showed decreased sugars as carbon is partitioned to terpene biosynthesis. This alternative C5 to C5 pathway was less disruptive of fixation by utilizing five carbon compounds from the pentose phosphate pool and diverting them to terpene synthesis, making a more efficient rechanneling of carbon directly from sugars to terpenes.

ACKNOWLEDGMENTS

Science rarely manifests as a single individual's endeavor alone. My study here at Texas A&M University has been testament to this, as my peers, collaborators, and professors have expanded my understanding and capabilities as a scientist. This dissertation seeks to compile some of that research, but a range of collaborations allowed for far more research to be accomplished during my time here in College Station. First, I would like to thank my committee members, Dr Herman Scholthof, Dr Mike Kolomiets, Dr Bhimu Patil, and Dr Joshua Yuan, who have been unbelievably accommodating of last minute signatures, meetings, and deadlines. I would like to especially thank Dr Joshua Yuan for recruiting me from undergraduate research to his lab and giving me the opportunity to work in his lab. I would also like to thank all the Chinese lab mates over the years which welcomed me into their homes and culture. I especially would like to thank my friend Dr Xin Wang for deepening my critical thinking and plenty of basketball at the rec center. I would also like to extend a special gratitude to my friends, Hu Cheng, and now Dr Cheng Zhao, who spent countless late hours with me extracting field samples, driving to and from North Carolina, and in general kept me sane throughout my time here. I also would like to thank the CEO of SynShark for the experience working in a biotech startup business operations as well as for the accommodations and hospitality of our farmer Pat Short who has become a personal friend. In getting me to this point, I would also like to thank Dr Neal Stewart who allowed me to be hands on in the lab and gain vital research experience prior to graduate study. Last, I would like to thank my family and friends back home who have supported me and encouraged me throughout my life and motivated me to make a better a world through my research.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a dissertation committee consisting of Professor Joshua S. Yuan [advisor, Department of Plant Pathology and Microbiology] and Professors Herman B Sholthof and Michael V Kolomiets of the Department of Plant Pathology and Microbiology, and Dr Bhimu Patil of the Department of Horticultural Sciences.

Engineering work for the material of G1 line tobacco were performed by Dr Joe Chappell's lab at the University of Kentucky. Line pT5 was generated by Dr Hong Ma and Dr YongkiYong Kim. Plasmid construction for Chapter III was completed along with Hu Cheng, who also did molecular characterization of RibB lines. Metabolomics analysis was completed using Metabolon® services.

All other work in this dissertation was completed by the student independently.

Funding for this work came from the U.S. Department of Energy, Advanced Research Projects Application in Energy (US-DOE, ARPA-E). Individual graduate study and research was funded by a departmental fellowship, as well as a teaching assistantship my final year.

NOMENCLATURE

MVA – Mevalonate (Mevalonic Acid) pathway

MEP - Methylerythritol 4-phosphate pathway, non-mevalonate pathway, (or Rohmer pathway)

IPP – Isopentenyl pyrophosphate

DMAPP –Dimethylallyl pyrophosphate

G3P – Glyceraldehyde 3-phosphate

DXP - 1-Deoxy-D-xylulose 5-phosphate

DXPS - 1-Deoxy-D-xylulose 5-phosphate synthase

RuBisCO – Ribulose-1,5-bisphosphate carboxylase/oxygenase

PDH – Pyruvate Dehydrogenase

DXR - 1-deoxy-D-xylulose 5-phosphate reductoisomerase (in some literature as IspC)

IspD – Isoprene Synthase D

MEcPP - 2-C-Methyl-D-erythritol-2,4-cyclopyrophosphate

IspG – Isoprene Synthase G

IspH – Isoprene Synthase H

HMBPP - (E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate

IDI - Isopentenyl-diphosphate delta isomerase

NADPH - Nicotinamide adenine dinucleotide phosphate, reduction cofactor

ATP – Adenosine tri-phosphate, energy carrier

TPS – Terpene Synthase

PHA – Polyhydroxyalkanoates, natural polyesters, bioplastic

C4 – photosynthesis using two cell types and channeling of malate, also called Hatch–Slack pathway

Calvin Cycle – light independent reactions of carbon fixation from atmospheric CO_2

SBPase - Sedoheptulose-bisphosphatase

Xu5P – Xylulose- 5-phosphate

Ru5P – Ribulose 5-phosphate

PEP - Phosphoenolpyruvic acid

PEPC – Phosphoenolpyruvate carboxylase

CAM - Crassulacean acid metabolism

MDH – Malate dehydrogenase

C3 – photosynthesis generating molecules of 3-PGA, a 3C molecule

ROS – Reactive Oxygen Species

plgg1 – plastidic glycolate glycerate transporter

GE – Genetically Engineered/Genetic Engineering

FPPS – Farnesyl pyrophosphate synthase

SQS – Squalene Synthase

G1 – Line designation of ctpFPPS+ctpSQS homozygous genotype

PGP - phosphoglycolate phosphatase

GO – Glycolate oxidase

CAT - Catalase

MS – Malate Synthase

ME – Malic Enzyme

ctp – Chloroplast Transit Peptide from RuBisCO small subunit

C_{14} - radiocarbon, is a radioactive isotope of carbon with an atomic nucleus containing 6 protons and 8 neutrons

WT – Wild Type, not genetically engineered, referring to line 1068 cultivar of *N. tabacum*

F6P – Fructose 6-phosphate

MS (Media) – Plant propagation basal medium originating from Murashige and Skoog

Lux – light measurement at a given area, measured as 1 lumen covering 1 square meter

pT5 – line designation of C2 photorespiratory bypass transformation into G1

HPLC – High Pressure Liquid Chromatography

GC/MS – Coupled Gas Chromatography/Mass Spectrometer

B10 – Growth chamber with cool temperature, moderate humidity, and low irradiance

B15 – Growth chamber with hot temperature, lower humidity, and high irradiance

(RP)/UPLC-MS/MS – Coupled Reverse-Phase/Ultra Performance Liquid Chromatography Tandem Mass Spectrometry

ESI – Electrospray Ionization

HILIC/UPLC-MS/MS - Hydrophilic Interaction chromatography/ Ultra Performance Liquid Chromatography Tandem Mass Spectrometry

QC – Quality Control

MS/MS – Tandem Mass Spectrometer

IAA – Indole-3-acetic acid (auxin)

GSH – Glutathione (antioxidant)

MAGs - Monoacylglycerols or monoglycerides

Welch's t-test – or unequal variances t-test, is a two-sample location test which is used to test the hypothesis that two populations have equal means

OAA - Oxaloacetic acid

FSTR – Designation to plants engineered with ctp:FPPS, SQS, T-oleosin, and RibB

FSSR – Designation of plants engineered with ctp:FPPS, SQS, SBPase, and RibB

IDT – Integrated DNA Technologies Company

DH5α - High efficintcey (line used heat shock compatible) *E. coli* competent cell line

10-β – High efficiency (line used electrocompetent) *E. coli* competent cell line

CaMV - Cauliflower mosaic virus promoter (CaMV 35S), denoted by genome origin

NeoR – Aminoglycoside phosphotransferase, confers neomycin resistance, same action as KanR

Poly-A – Polyadenylation, the addition of a poly(A) tail to a messenger RNA, disassociation/termination sequence in DNA

Pcv – Truncated Cassava vein mosaic virus promoter

Tnos - Termination sequence of the nopaline synthase gene, isolated from *Agrobacterium tumefaciens*

p35S – Same as CaMV, denoted by the coefficient of sedimentation of the viral transcript driven by this promoter

T-oleosin – Truncated oleosin, structural protein, makes up oil droplet

T-DNA – Transfer-DNA, DNA to be moved from *A. tumefaciens* binary plasmid to plant cell, flanked by left and right border sequences

GV1301 – Disarmed lab strain of *A. tumafaciens*

PCR – Polymerase Chain Reaction

YEP – Yeast Extract and Bacto-Peptone media for *A. tumafaciens*

OD – Optical Density

BAP - 6-Benzylaminopurine, synthetic cytokinin plant growth regulator

NAA - N-Acetylaspartic acid, synthetic auxin plant growth regulator

T5 – Tube size for commonly used fluorescent lighting tubes in plant propagation

HPS – High Pressure Sodium light bulb

T_0 – The initial transformants, plants growing from individual events of genetic engineering

T_1 – The seed of T_0 line, the first generation from seed after genetic engineering

DNA – Deoxyribonucleic Acid

mRNA – Messenger ribonucleic acid, transcripts, products of gene transcription

(RT)-PCR – Reverse transcriptase PCR, synthesizes DNA from RNA, Amplicons from transcripts

RuBP – Ribulose Bisphosphate, 5C needed in RuBisCO reactions with CO_2

HMG - 3-hydroxy-3-methylglutarate

HMG-CoA - 3-hydroxy-3-methylglutarate coenzyme-A

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

I. INTRODUCTION AND LITERATURE REVIEW

Terpenes and Terpene Biosynthesis

The production, extraction, and consumption of terpenes has existed among mankind for ages[1]. The application of terpenes are broad, as they are used as medicines, biofuels, specialty chemicals, nutraceuticals, therapeutics, fragrances, etc.[1, 2]. Many of these have highly diverse utilities; limonene for example is used in the food and fragrance industry as well as being used as a cancer preventative and a fuel stabilizer[3]. Others, such as squalene and taxadiene have high economic values[4]. Terpenes and terpenoids, terpenes with additional functional groups, make up one of the largest groups of natural products. Sometimes referred to as Isoprenoids from their common 5-carbon isoprene structured backbone[5], terpenes have diverse structure and therefore function making them useful bioproducts such as pharmaceuticals or nutraceuticals, fuel additives and stabilizers, topical products and fragrances, and some yet undiscovered uses.

Terpene biosynthesis involves two pathways, the earlier elucidated MVA (Mevalonate) pathway, which exists in eukaryotic cytosol and archaea organisms, and the later discovered MEP (Methylerythritol 4-phosphate) pathway present in most prokaryotes, green algae, and plants[6, 7]. In the MVA pathway, acetyl-CoA is used to produce IPP (Isopentyl Diphosphate/Pyrophosphate) and DMAPP (Dimethylallyl Diphosphate/Pyrophosphate), with Mevalonic Acid as an intermediate[8]. The MEP (Methylerythritol 4-phosphate) pathway begins from the condensation of pyruvate and G3P (Glyceraldehyde-3-phosphate/GAP) into DXP (1-deoxy-D-xylulose 5-phosphate) by DXPS within the chloroplast lumen. This reaction yields CO_2, which acts to increase CO_2 near RuBisCO activity. DXPS specifically catalyzes the thiamine pyrophosphate- and Mg (II)-dependent conjugation of pyruvate and G3P to form DXP and CO_2[9]. In the absence of G3P,

DXPS could potentially catalyze the condensation of two pyruvate molecules to form acetolactate [10]. In the chloroplast, endogenous PDH (pyruvate dehydrogenase) activity also competes with DXPS for pyruvate, which yields acetyl-CoA and CO_2, decreasing pyruvate pool but also increasing CO_2 concentration in the chloroplast. In the MEP pathway, DXP is then reductively isomerized to MEP by DXR[11]. IspD, IspE and IspF (Isoprene Synthases) catalyze the activation of MEP and then cyclization to form MEcPP (2-C-Methyl-D-erythritol-2,4-cyclopyrophosphate). IspG catalyzes the reductive dehydration of MEcPP, and is less understood than other enzymes in the pathway[8]. It is an iron-sulfur [4Fe-4S] cluster enzyme and has been shown to have activities which are orders of magnitude lower than other MEP enzymes, suggesting a limiting step in terpene biosynthesis. IspH catalyzes the reductive dehydration of HMBPP ((E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate) from IspG to produce DMAPP and IPP. IspH produces DMAPP and IPP at a ratio of 5:1, however IDI (Isopentenyl diphosphate isomerase) is able to perform a reversible reaction of DMAPP to IPP[8]. Accumulation of IPP acts as a product feedback inhibition to DXPS. The plastid MEP pathway produces monoterpenes (C_{10}) and duplications thereof, diterpenes (C_{20}) and tetraterpenes (C_{40}). The cytosolic MVA pathway produces sesquiterpenes (C_{15}) and triterpenes (C_{30}). Plants contain both fully functioning plastid MEP and cytoplasmic MVA, whereas green algae and cyanobacteria predominately only contain the MEP[2]. This makes plants unique for terpene production in terms of engineering. The MEP requires reducing power and energy in the form of NADPH and ATP from light reactions[12]. Along with the pathways direct coupling to G3P, the MEP is highly regulated by photosynthesis rate, and limitations to fixation efficiency by available reduction potential. Carbon for the MEP starts from G3P, which makes terpene production a carbon sink for photosynthesis, which can be enhanced with various strategies.

Plants utilize terpenes for various uses as well, therefore their biosynthesis is tightly regulated both temporally and spatially[13]. From an evolutionary perspective, terpene biosynthesis is largely utilized in signaling and plant defenses against pathogens and insects, driving a wide range of products and enzyme efficacy within the MEP and MVA pathways as well as TPS (Terpene Synthases)[14]. Furthermore, terpenes are hydrocarbons which are limited by primary metabolism and carbon allocation[2]. Because of the range of products capable of being synthesized from terpene biosynthesis, many studies seeking to increase bioproduct yield have been conducted [2, 15-20]. Both heterotrophic and photosynthetic organisms have been engineered, however photosynthetic organisms offer the ability to synthesize hydrocarbons from atmospheric CO_2 by carbon fixation. This could lead to many new product synthesis routes from excess atmospheric CO_2 moving away from fossil fuels and recycling their current emissions. Furthermore, the use of carbon fixation directed into hydrocarbon bioproducts offers a closed energy loop between human use sustainable production schemes, whereby the emissions of burning hydrocarbon based fuels produces CO_2 to be captured and returned to hydrocarbon again using light as a catalyst.

The Application of Metabolic Engineering

Metabolic engineering allows synthesis of novel bioproducts through utilization of common metabolites as intermediates. The application of synthetic biology to produce compounds previously limited by source offers sustainable routes to pharmaceuticals, high value fuels, and industry or market demanded commodities. One promising avenue of metabolic engineering is the synthesis of hydrocarbons, the base of fuels and plastics. Many systems using heterotrophic microbial biotechnology have yielded useable natural gases such as methane[21], and bioplastics such as PHA[22]. While many hydrocarbons have been synthesized using biotechnology in

heterotrophic systems, production in a photoautotrophic platform offers solar driven biological conversion of atmospheric CO_2 to hydrocarbons directly in the form of terpenes[2]. Many plants produce a wide array of terpenes and terpene derivatives, but accumulating biofuel from plants is less feasible than a microorganism. Plants do however offer a unique production system due to multiple compartmentalizations and the existence of both terpene precursor synthesis pathways, MVA and MEP. Plants are also capable of large chain hydrocarbon synthesis and many are prolific oil producers[2, 23]. Even in their natural titer many terpene bioproducts are extracted and marketed profitably[1].

The production of some terpene bioproducts comes at an economic or environmental cost. This is certainly the case with the triterpene, squalene. Procurement of squalene traditionally comes from two sources. The primary source is from the liver of deep water sharks which are delivered and left to die, a highly unsustainable environmental problem when one considers the growing demand for squalene and decreasingly abundant ocean. Although squalene primarily comes from deep water shark livers, which is where the name comes from (Squalus is the genus for sharks), due to market demand and unsustainable procurement practices, some squalene is being produced from plant and alternative sources[24]. Research into producing squalene in yeast (*S. cerevisiae*) yielded a lower value product, squalane, while consuming plant grown sugars from sugar cane in a fermentation system to first produce farnesene. Farnesene is then dimerized to isosqualene and then hydrogenated to squalane[25]. This hydrogenation product of squalene lacks the double bonds and is thus fully saturated, making squalane lose the UV protection and ROS (reactive oxygen species) scavenging capacity of squalene. Other attempts to use plant based methods from olive oil, rice and sugar cane have been derived. However extracts from olive oil and other plant based cooking oils, which already bring a premium price, are rendered unusable

after the removal of squalene. This makes production from oil crops less economically viable as an option for squalene bioproduction. Similarly rice, (rice bran), is being used for the procurement of squalene, however most of the material sold and advertised is squalane. This along with rice being a staple of much of the world's population, makes rice less likely to replace shark delivering as the market demand continues to increase. Sugar cane has been a target both for making squalene from fermentation and engineered yeast, as well as developing more direct routes to channel the excess carbon put into sucrose and starch production directly into terpene biosynthesis. While engineered biosynthesis of squalene was not achieved in sugarcane, the possibility of utilizing C4 organisms and carbon concentrating mechanisms remains a possibility in increasing fixation rates required for high hydrocarbon bioproduct synthesis.

In this thesis, we used squalene as a model compound to study how to rewire photosynthetic output to produce higher level of terpene. The synergy of 'source' and 'sink' will enhance terpene productivity and provide a practical solution for producing a high value compound with an increasing market, squalene.

Limitations in Terpene Biosynthesis

Despite gains in yields by various engineering teams, terpene biosynthesis is largely limited in terms of the yield and productivity that can be achieved for engineering purposes *in planta*. In the MVA, supplying carbon and energy for biosynthesis largely relies on catabolism[8]. In the MEP carbon is borrowed from the Calvin Cycle as G3P, along with reducing power and energy, limiting the carbon flux through the MEP as it is allocated to primary metabolism first[8]. This allocation of carbon, reducing power, and energy all are highly dependent on photosynthesis, and therefore greatly impacted by environmental conditions. The MEP is further regulated internally,

both by product feedback inhibition (IPP and DMAPP to DXPS), as well as possible limited rate steps through Isoprene synthases (IspD-H)[8]. Engineering terpene bioproducts in photosynthetic organisms shows promising opportunities, but has many remaining challenges. Primary challenges with producing elevated yields of terpene bioproducts other than a strong constitutive sink remain in (1) The carbon fixation efficiency and allocation to the MEP pathway, (2) carbon flux and productivity of the MEP pathway itself, and (3) along with a strong sink, an accumulation strategy must be implored.

One measurable gain to all bioproduction remains carbon fixation efficiency. Increased carbon in any crop allows for higher yield, increased vigor, and larger biomass. The ability to produce hydrocarbons from atmospheric CO_2 is no different. Efforts have been made with some success in increasing biomass and vigor through the introduction of high efficiency SBPase (Sedoheptulose-bisphosphatase)[26-29], which catalyzes the removal of a phosphate group from sedoheptulose 1,7-bisphosphate to produce sedoheptulose 7-phosphate. This shortens the regeneration step in the Calvin Benson cycle helping to produce 5C precursors Xu5P and Ru5P from a 7C sugar phosphate (sedoheptulose-bisphosphatase) and a 3C sugar phosphate (G3P). In plants engineered with the photorespiratory bypass, terpene sink, and transgenic SBPase from cyanobacteria, the highest yield observed to date at over 7mg/g FW (fresh weight) squalene per fresh weight tobacco occurred. This was approximately 3 fold more than the photorespiratory bypass with terpene sink line (2.7mg/g FW), and 10 fold the yield by terpene sink alone (600ug/g FW). These lines also exhibited elongated lance shaped leaves, extreme variability, and low tolerance to high light and heat. The highest observed squalene yield occurred in January, when the greenhouse temperature and light were decreased from spring and summer greenhouse propagation. These plants were tested in an outdoor pilot program field plot study where these

lines showed slow growth and low survivability compared to WT and other engineered lines. One possible explanation is that SBPase competes for G3P with the MEP pathway DXP enzyme, and sugar biosynthesis. Furthermore regulation of SBPase by endogenous factors and variable feedback loops made consistency in yield very irregular[26].

One of the major limits in carbon fixation efficiency relies in the efficacy of RuBisCO itself and its regulation, namely by RuBisCO activase[30]. Nature has solved the issue of RuBisCO acting as both a carboxylase and oxygenase in environments when increased light intensity or heat would cause increased oxygenase activity during peak light hours. Increasing photon density in the form of solar radiation increases light intensity as well as heat. At higher temperatures RuBisCO acts increasingly as an oxygenase, and at temperatures above 35°C *in vivo* becomes deactivated[31]. This regulation is done by RuBisCO activase, as isolated RuBisCO enzyme in vitro has shown catalytic rate to increase at 42°C and >50°C [31]. When temperature rises, RuBisCo deactivation rate exceeds that of the rate RuBisCO activase can promote activation[31]. This reduces the RuBisCO activity by a regulator which is independent of CO_2 concentration. RuBisCO activase functions to mitigate oxygenase reactions by reducing RuBisCO activity when the elevated temperatures loosen the kinetic sites of RuBisCO[31]. Itself, RuBisCO activity is kinetically regulated by temperature and activity regulated by relative concentrations of CO_2 and O_2 [31]. Coevolution of Kranz leaf anatomy and C4 carbon fixation, as in maize, operate to concentrate CO_2 near RuBisCO reactions. This increases the CO_2 concentration within the bundle sheath cell by using PEP (phosphoenolpyruvate) to shuttle a carbon as a 4-carbon acid by PEPC (phosphoenolpyruvate carboxylase) from the mesophyll. From the 4-carbon intermediate, CO_2 is evolved in the bundle sheath cell, exporting PEP back to the mesophyll[32]. Plants in arid dessert ecosystems must adapt further. In addition to high heat, they receive less water if any for long

periods of time. Named after the taxonomic family also referred to as stonecrops or succulents, Crassulacean Acid Metabolism (CAM) collects CO_2 at night where carbo-anhydrase produces bicarbonate which PEPC uses to carboxylate PEP and is then reduced by MDH to form malate[33]. This is similar to the transporting system and use of bicarbonate, PEP and PEPC to move CO_2, but using a temporal separation using malate and malic acid. Malate is transported into and stored in the vacuole as malic acid, and transported back out and catabolized to CO_2 for carbon fixation and pyruvate[33]. RuBisCo itself also generates some pyruvate within the chloroplast as a minor by-product of catalysis, generating pyruvate at a rate of 0.7% per ribulose bisphosphate and increasing with temperature[31, 34]. Both C4 and CAM photosynthesis offer increased fixation efficiency over C3 plants in dry or arid environments respectively, however they may be limited for production of bioproducts due to[35] lack of biomass accumulation or titer of prospective bioproducts[33]. These organisms do however offer sources for enzymes with functional applications in engineered alternative carbon fixation and allocation schemes to bioproducts. Partial strategies exist in nature for accumulating CO_2 near RuBisCO reactions without physiological evolution in lower order organisms such as bicarbonate transporters, although this produces a heterotrophic system based on the supplement of a carbon substrate in the form of the bicarbonate[2]. In our strategy, we overcame the challenge by scavenging the oxidative products from RuBisCo when it is non-specifically binding with oxygen.

Methods to create a more efficient RuBisCO enzyme have been attempted and continue, including remodeling of its subunits, directed evolutionary studies, and mutagenesis work[36-38]. While some gains have been made, the duality of this enzyme and dynamic nature of its carboxylase/oxygenase activities have shown to have regulatory rolls[30, 38]. Photorespiration dissipates reaction energy driven by light reactions into an alternative route which also removes

O_2 from the RuBisCO reaction site[39]. Excess energy and increased O_2 concentrations can lead to photooxidation, either directly or via production of ROS after photoinhibition[39]. Energy of photorespiration itself has multiple fates driven by energy dissipation modes, ATP-producing or O_2-producing[39]. Through production of ATP and NADPH, photorespiration acts to recover energy lost, while also balancing energy and reductant ratios[39]. Removal of or inhibition of many photorespiration elements leads to weakened phenotypes and typically death[40]. Similarly, many dissipation modes rely on primary metabolism elements as essential elements as well, such as catalases, transporters, and energy carriers[39]. It is however possible to shorten the route of carbon from photorespiration back to the Calvin Benson cycle or to direct it to alternative metabolic needs. So long as the photorespiratory pathway is not removed, it will compete with a synthetic downstream pathway. Attempts to reroute the product of RuBisCO oxygenase reactions, p-glycolate, have had success in increasing the 2C compound to terpene production[35, 41]. Blocking the *plgg1* transporter in tandem with rerouting glycolate has shown to have significant biomass gains, showing promise of exploitation by metabolic engineering as a source of additional carbon for bioproduct synthesis [35].

Alternatively, routes to increase the carbon allocation from fixation or heterotrophic sources to terpene bioproducts has been explored vastly, with many companies currently on market producing from GE organisms. Given all terpenoids come from the precursors of DMAPP and IPP, the synthesis of new drugs and fuels from photosynthesis will rely on the ability to move fixed carbon to these isoprenoid precursors. This requires optimization of the MEP pathway itself. Characterization of its enzymes and kinetic steps has elucidated a few challenges and possible bottle necks. The carbon flux to the MEP pathway is limited as mentioned above by light reactions for energy and reductant but also by availability of G3P (assuming pyruvate is not limited) as a

substrate for the first committed step in terpene biosynthesis. This relies on DXPS to compete with primary metabolism enzymes for both limited G3P and commonly used pyruvate. Increasing fidelity and enzyme titer of DXPS have shown to increase carbon flux to downstream terpene synthesis engineered sink [2, 42]. Terpene synthases (TPS) are typically kinetically slow enzymes, with exception to some volatile TPS, and offer a further step of engineering to produce faster enzymes as well as new enzymes with synthetic biology[43, 44].

Accumulation of terpene bioproducts also remains a challenge. Utilized as metabolic intermediates or evolved as volatiles, many terpenes do not naturally accumulate, even if present in high titer or synthesized at a high rate[13, 43]. Spatial partitioning, separated from downstream enzymes or engineered into systems without a downstream metabolism, has allowed for increased accumulation of some terpene bioproducts[13]. In plants, terpenes can be stored and secreted into specialized physiological structures for tertiary metabolites including trichromes, sheath cells, and even vascular tissues[13]. For production in plantae however, excretion or emission of the bioproduct make capture less feasible, than in a microbial bioreactor type system. Engineered terpene production has been collected from volatile emission in algae and cyanobacteria[2], and from trichromes of tobacco[45]. Also, in tobacco engineered with a squalene synthesis sink within the chloroplast was able to accumulate squalene. Spatial separation from endogenous downstream cytoplasmic phytosterol production allowed extractable titers of the target bioproduct[45]. The engineering of a synthetic lipid vesicle to encapsulate squalene leaked from the chloroplast membrane into tobacco lines previously engineered with the terpene sink further increased squalene titer[46]. Endogenous terpene/terpenoid bioproducts typically do not accumulate even when synthesis rate is high for three reasons; (1) The bioproduct is an intermediate metabolite and thus consumed, (2) The resultant compounds can be relocated to other intracellular compartment

and consumed or the bioproduct is volatile allowing for emission/evolution of the bioproduct to the environment, or (3) The bioproduct undergoes reduction or oxidation reaction such as those in plastids or the vacuole, similar to the first reason this yields another product than the desired bioproduct. For this reason exogenous terpene/terpenoid biosynthesis, as such created by synthetic biology, allows for easier accumulation and subsequent extraction. The synthesis of a metabolite within an organism with neither downstream metabolism nor catabolism for the product will accumulate the product. This can lead to other issues however such as cytotoxicity and inhibition of other pathways, however most engineering efforts have not reached these levels as desired bioproduct accumulation still remains a challenge.

Construction of a Novel Terpene Sink

The increased yield of the terpene bioproduct squalene was produced creating a strong constitutive terpene carbon sink downstream of the MEP pathway. This was constructed using FPPS (Farnesyl pyrophosphate synthase) and SQS (Squalene synthase), which are typically cytosolic enzymes, and attaching a chloroplast transit peptide (ctp) to be functional in the chloroplast stroma[45]. The resulting genotype from transformation into Nicotiana tabacum L. 1068 was designated G1. This allowed terpene precursors, IPP and DMAPP, from the MEP to produce squalene within the chloroplast[45]. Because squalene is a triterpene, but synthesized in the plastid, it is partitioned from downstream cytosolic phytosterol biosynthesis, allowing squalene to accumulate[46]. This novel sink allows for furthered study into engineering the MEP pathway and carbon flux from primary metabolism to terpene biosynthesis. Increased bioproduct titer with a strong terpene carbon sink still relies on carbon allocation both to and through the MEP pathway. From the novel sink provided by G1, upstream modifications to carbon metabolism could be constructed using the terpene synthesis carbon sink as a metric of gain for carbon allocation from

primary metabolism to terpene bioproduct yield. From conception of the FPPS and SQS sink, permutations of MEP enzymes were tested along with carbon rerouting to DXPS (pRD1-4) seeking to increase accumulation of squalene. Modest gains were made by the addition of DXPS, the first step in the MEP pathway[45], but reaching yields feasible for production at a commercialization scale were folds away. The MEP is not just an energy and reductant deficient pathway, but also a carbon deficient pathway[47, 48]. In order to produce more terpene bioproduct more carbon has to be allocated to the MEP. In a push-pull scenario of the MEP the strong constitutive sink pulls metabolites in one direction, but is limited by what can be pulled through the MEP. The pathway is always limited by carbon into the pathway and flux through the pathway. Both are targets for engineering and study moving forward in synthetic biology.

II. CONSTRUCTION AND FUNCTIONAL VALIDATION OF THE C2 PHOTORESPIRATORY BYPASS PATHWAY

Introduction

The MEP pathway has many limitations in the capacity to synthesize terpene bioproducts at economically feasible titers. One large limitation is the carbon allocation to the MEP pathway itself. For this purpose, a C2 Photorespiratory Bypass pathway was transformed into the engineered G1 line of *N. tabacum* containing the chloroplast transit peptide fused FPPS and SQS coding sequences. Validation of modeling performed by squalene yield, photosynthetic rate, and C_{14}-labeling experiments show rerouted carbon from the photorespiratory byproduct glycolate incorporated into the MEP pathway and squalene biosynthesis.

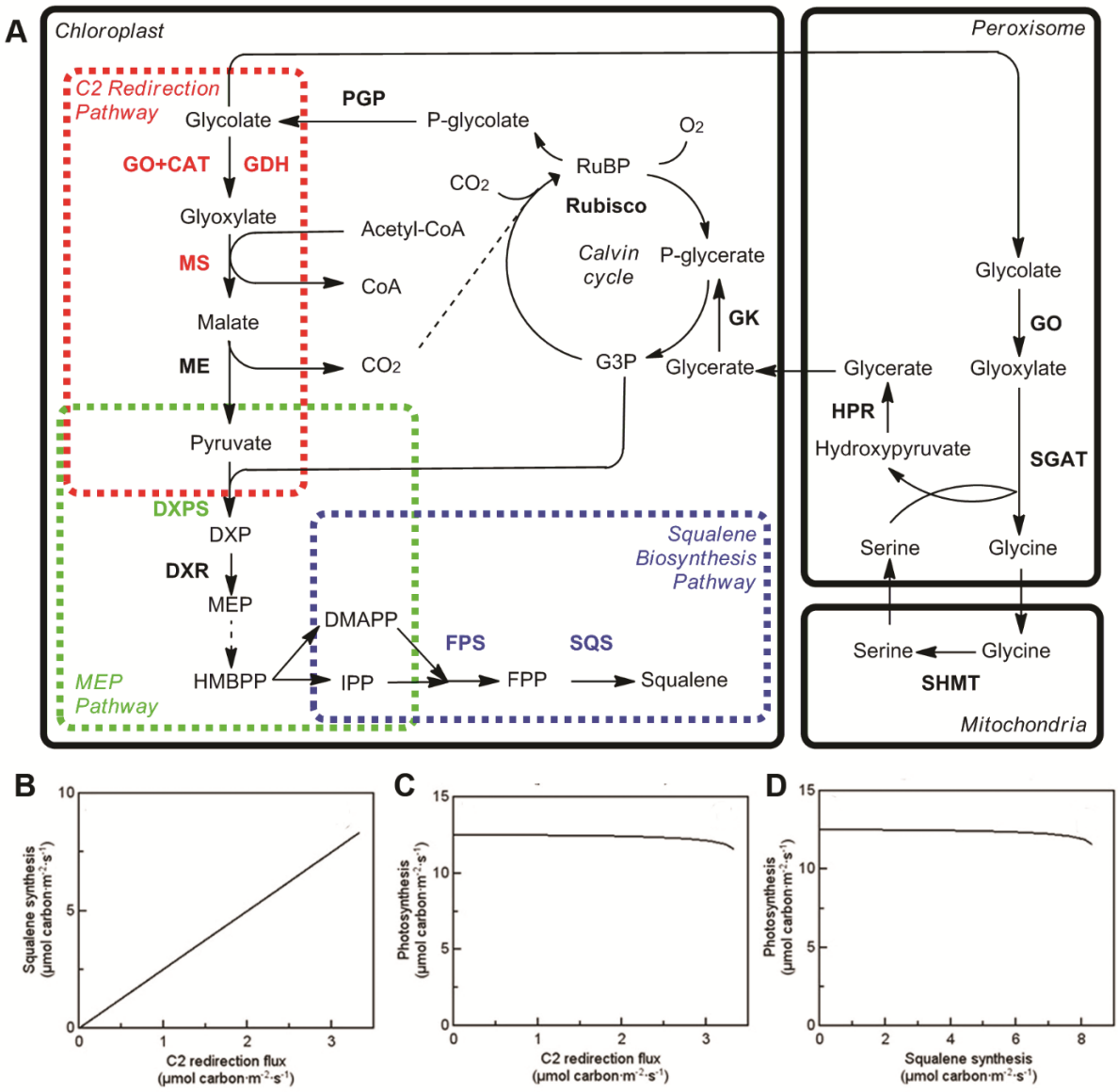
In order to supply the MEP with increased carbon flux, a pathway to reroute a portion of photorespiratory byproducts to terpene synthesis was engineered as designed and shown in **Figure 1A**. Instead of rerouting glycolate through the peroxisome and mitochondria, the bypass seeks to partition part of the glycolate into malate synthesis, followed by cleaving malate into needed pyruvate and CO_2. Although some rerouted carbon is lost here and in the condensation of pyruvate and G3P, the evolution of CO_2 works to recover carbon lost by causing increase in CO_2 concentration near RuBisCO activity, producing a favorable fixation environment. The immediate product of photorespiration yields p-glycolate which is rapidly converted into glycolate by PGP (phosphoglycolate phosphatase) where the C2 photorespiratory bypass begins. Glycolate is converted to glyoxolate by GO (glycolate oxidase) along with CAT (catalase) to remove the H_2O_2 generated by the GO reaction. Next, MS (malate synthase) catalyzes the reaction of glyoxolate and acyl-CoA to malate. Endogenous ME (malic enzyme) produces pyruvate from malate, evolving a CO_2 in the process. From here the initial committed step into the MEP is satisfied when pyruvate

and a molecule of G3P from the Calvin Benson cycle are condensed, forming DXP (1-Deoxy-D-xylulose 5-phosphate) by DXPS (DXP synthase). DXP is then converted to MEP (2-C-Methyl-D-erythritol 4-phosphate) by DXR (DXP reductoisomerase) and reduction energy from NADPH. MEP is then further converted to DMAPP and IPP by Isoprene synthases (IspD through IspH). Terpene synthases (TPS) then utilize chloroplast DMAPP and IPP to produce squalene via the engineered terpene synthesis carbon sink created by ctpFPPS and ctpSQS [49]. Modeling for this alternative pathway shows no change in fixation rate, while showing incorporation of redirected carbon through the MEP and into terpene production (**Figure 1C&D**). The modeling also shows that increased levels of redirected C2 (glycolate) correlate to a linear increase in squalene yield (**Figure 1B**).

Figure 1. Pathway Design and Modeling of C2 Photorespiratory Bypass

Pathway design and modeling result. (**A**) A schematic outline of pathway design in this study. The C2 redirection pathway, the MEP pathway and the squalene biosynthesis pathway were coupled to produce squalene in tobacco chloroplast. The enzymes were indicated by bold font. The colored metabolic steps were engineered in this study. CAT, catalase; DXPS, 1-deoxy-D-xylulose 5-phosphate synthase; DXR, 1-deoxy-D-xylulose 5-phosphate reductoisomerase; FPS, farnesyl diphosphate synthase; GDH, glycolate dehydrogenase; GK, glycerate kinase; GO, glycolate oxidase; HPR, hydroxypyruvate; ME, malic enzyme; MS, malate synthase; PDH, pyruvate dehydrogenase; PGP, phosphoglycolate phosphatase; SGAT, serine-glutamate aminotransferase; SHMT, serine hydroxymethyl transferase; SQS, squalene synthase; DMAPP, dimethylallyl diphosphate; DXP, 1-deoxy-D-xylulose 5-phosphate; FPP, farnesyl diphosphate; HMBPP, 4-hydroxy-3-methylbut-2-enyl-diphosphate; IPP, isopentenyl diphosphate; MEP, 2-C-methyl-D-erythritol 4-phosphate; G3P, glyceraldehyde 3-phosphate; RuBP, ribulose 1,5-biphosphate. (**B**) The modeling result of terpene production rate in response to C2 redirection. (**C**) The modeling result of photosynthetic carbon assimilation rate in response to C2 redirection. (**D**) The modeling result of photosynthetic carbon assimilation rate in response to squalene production in C2 redirection plants.

Figure 1.


A
Chloroplast
Peroxisome
C2 Redirection Pathway
PGP
Glycolate
P-glycolate
O2
RuBP
CO2
Rubisco
GO+CAT
GDH
Glyoxylate
Acetyl-CoA
Calvin cycle
P-glycerate
MS
CoA
Glycolate
GO
Malate
GK
ME
CO2
G3P
Glycerate
Glycerate
Glyoxylate
HPR
Hydroxypyruvate
SGAT
Pyruvate
DXPS
DXP
Squalene Biosynthesis Pathway
Serine
Glycine
DXR
MEP
DMAPP
FPS
SQS
Serine
Glycine
SHMT
MEP Pathway
HMBPP
IPP
FPP
Squalene
Mitochondria
B
Squalene synthesis (μmol carbon·m-2·s-1)
C2 redirection flux (μmol carbon·m-2·s-1)
C
Photosynthesis (μmol carbon·m-2·s-1)
C2 redirection flux (μmol carbon·m-2·s-1)
D
Photosynthesis (μmol carbon·m-2·s-1)
Squalene synthesis (μmol carbon·m-2·s-1)

Results

Even at efficient photosynthetic rates photorespiration occurs at relatively high frequencies, leading the team to believe that recovery of a portion of photorespiratory carbon and redirecting it to terpene biosynthesis will increase bioproduct yields. Observationally, the modeling has been confirmed, functionally increasing squalene by rerouting carbon. Modeling confirmed by C_{14}-labeled glycolate feeding assays demonstrated incorporation of labeled carbon in both malate and squalene, labeled feeding results are shown in **Figure 2**. The C2 photorespiratory bypass effectively rerouted a portion of photorespiration byproduct carbon to the terpene bioproduct carbon sink. The highest squalene accumulation in fully mature leaves from 3-month-old C2 redirection lines was 2.7 mg/g FW (fresh weight), about four-fold higher than the approximately 600 μg/g FW in the ctpFPPS and ctpSQS over-expressing lines. Also observed was high variability between samples with respect to squalene yield, having increased variation between samples of the C2 redirection plants without a change in photosynthesis (**Figure 3**). Variance between samples of different time points were attributed to the plant and leaf samples' age and health.

The C2 photorespiratory bypass plants showed significant variation in squalene yield when grown in different environments, which suggests a large environmental effect on terpene yield. Plants were grown in multiple greenhouse locations with varying amount of artificial additive light, temperatures, and humidity. At the same time the technology was licensed to SynShark LLC to conduct pilot field studies to determine feasibility and yield of the squalene producing engineered tobacco in Greensboro, NC. Traditional field grown engineered tobacco recovered their typical burley physiology, when compared to the etiolated and pale plants in the greenhouse.

Figure 2. C_{14}-glycolate Feeding Assay Incorporation into Malate and Squalene

C_{14}-glycolate feeding assay. (**A**) A schematic outline of the entire C2 redirection pathway design. Tobacco leaf discs were fed with C_{14}-glycolate, and the radioactivity incorporation into malate and squalene were examined. (**B-E**) The incorporation of radioactivity into malate. Two individual T_1 generation transgenic plants of pRD3 and pRD4 (C2 photorespiratory bypass) lines were analyzed. The means and standard deviation of three independent experiments were presented in the figure. (**F-I**) The incorporation of radioactivity into squalene. Two individual T1 generation transgenic plants of pRD3 and pRD4 lines were analyzed. The means and standard deviation of three independent experiments were presented in the figure.

Figure 2.

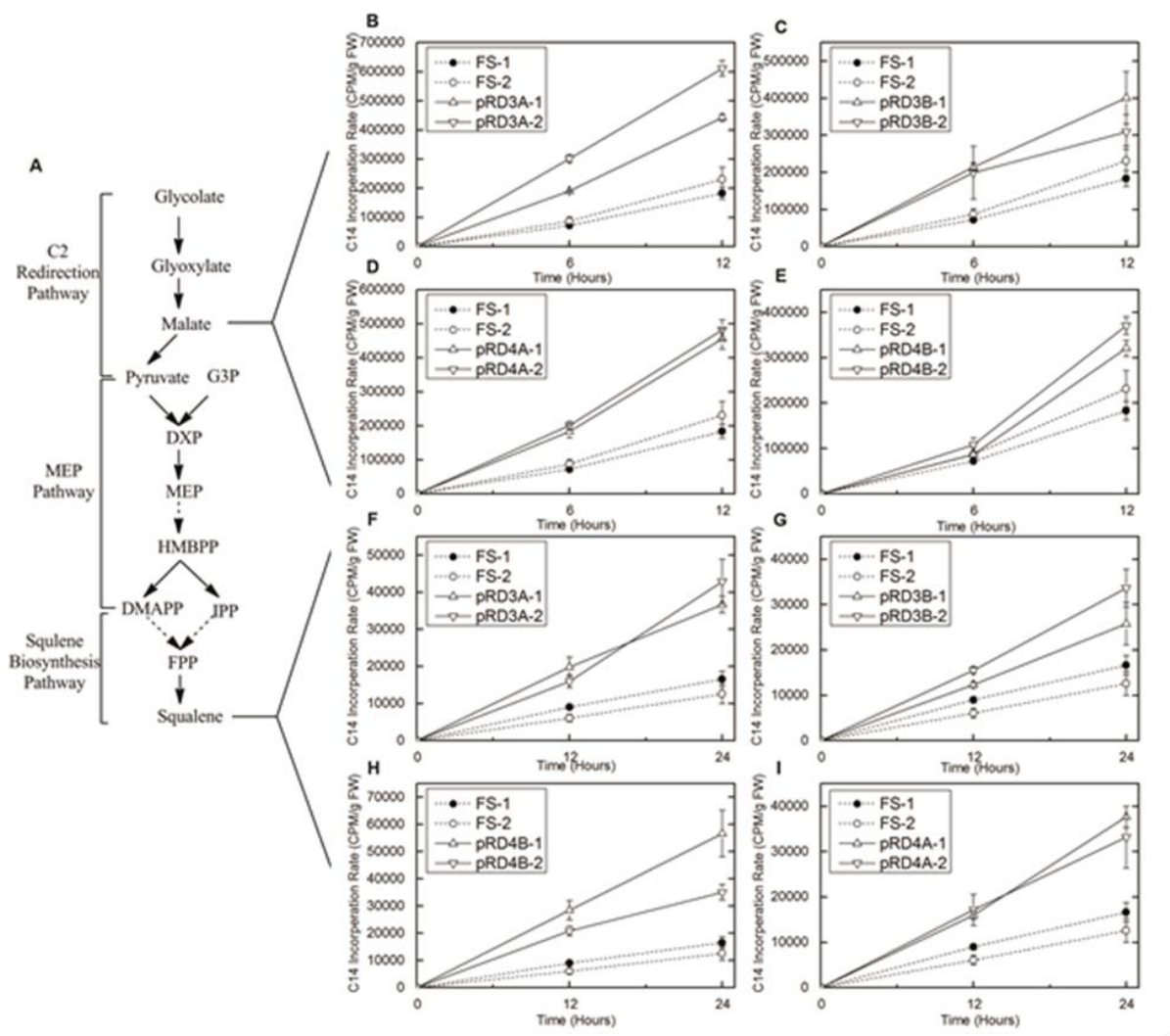
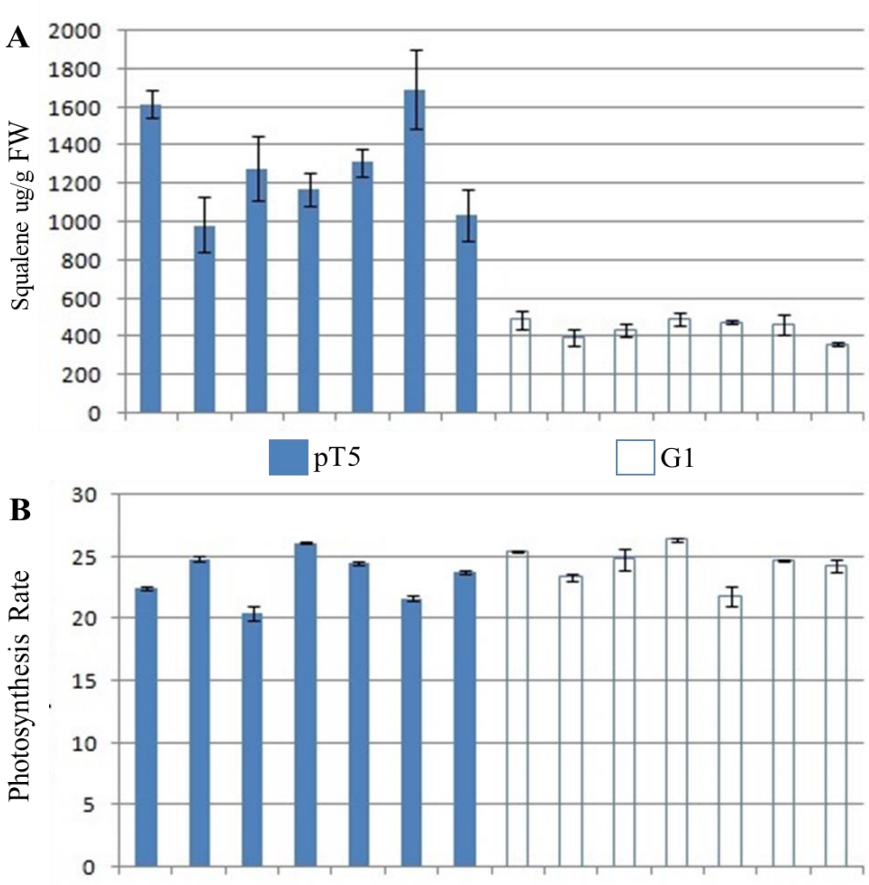

A
Glycolate
Glyoxylate
Malate
Pyruvate
G3P
DXP
MEP
HMBPP
DMAPP
IPP
FPP
Squalene
C2 Redirection Pathway
MEP Pathway
Squlene Biosynthesis Pathway
B
C
D
E
F
G
H
I
FS-1
FS-2
pRD3A-1
pRD3A-2
pRD3B-1
pRD3B-2
pRD4A-1
pRD4A-2
pRD4B-1
pRD4B-2
C14 Incorperation Rate (CPM/g FW)
Time (Hours)

Figure 3. Squalene Yield and Photosynthesis Rate of pT5 and G1

Quantification of squalene yield (**A**) determined by GC/MS, and photosynthesis rate (**B**) determined by LiCor® LI-6400, of C2 photorespiratory bypass plants (pT5) and G1 plants (chloroplast FPPS & SQS) grown in College Station field plot. (**A**) Shows a high intragroup variation for the bypass line among individuals, along with variation in technical replications. The wide yield range among all measurements within the group when compared to variation observed in the chloroplast FPPS and SQS only lines suggests variation is introduced by the bypass design. (**B**) No significant variation is observed in photosynthesis rates between groups, showing additional squalene yield is not caused by increased fixation rate.

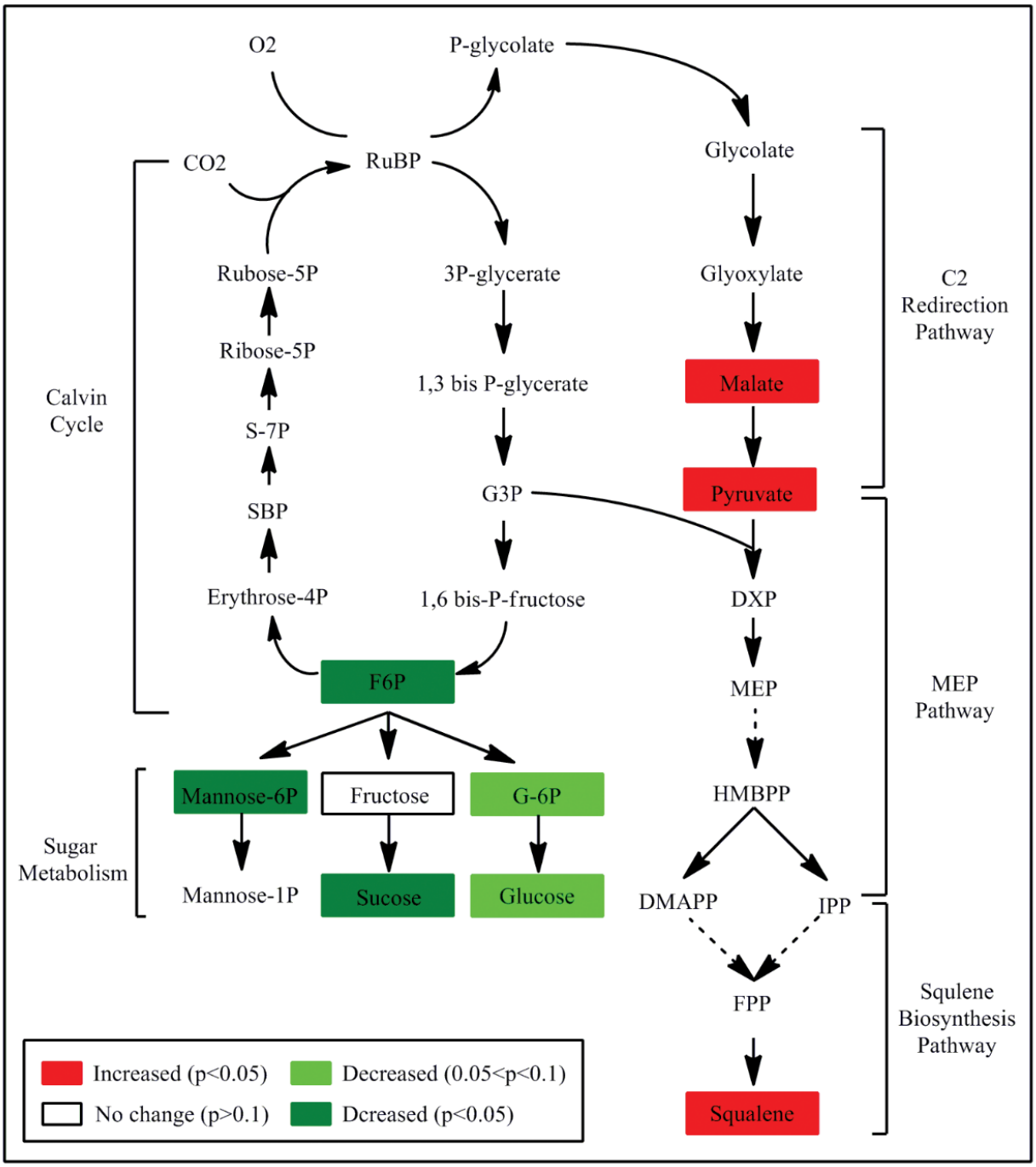


Similar to the greenhouse experiments, plants with C2 photorespiratory bypass yielded higher squalene than those with the chloroplast functional FPPS/SQS sink alone. Also worth noting was a small but significant reduction in biomass from the wild type (WT) burley background (L. 1068) to all transgenic lines, which could also be attributed to row effect and propagation of transgenic seed on selection media. Without a significant increase in photosynthesis between lines (WT, C2 redirection bypass, and FPPS/SQS sink plants) the relative accumulation of terpene bioproduct can be described by carbon allocation from primary metabolism to secondary and tertiary metabolism, guided by metabolic engineering (**Figure 4**). To reallocate carbon to and through the MEP pathway requires redirection from sugar biosynthesis in the form of G3P allocated to DXPS along with pyruvate largely derived from malate. The reduction in global sugars especially in F6P (Fructose 6-phosphate) and sucrose observed by metabolomics analysis show successful diversion of primary metabolism carbon into squalene biosynthesis in C2 photorespiratory bypass plants at the cost of sugar biosynthesis. The diversion from primary metabolism is indicative of a strong carbon sink into squalene synthesis. While the C2 photorespiratory bypass does increase yields, its intrinsic connection to light reactions and the Calvin Benson cycle allow for high variability with concern to predictable yield when the growth environment is altered, as well as typical temporal and spatial yield variations expected with phyto-bioproducts such as biomass age and health. To better understand the advantages and limitations in the photorespiratory bypass system as it responds to environment and in which ways the environment mitigates or exacerbates genetic variances of these plants compared to terpene sink only plants, a controlled environment study is required. In this way quantification of metabolic variance will highlight differences between genetic and environmental impacts on global metabolism and understanding of contributing factors to bioproduct yield gains.

Figure 4. Carbon Flux Model of C2 Pathway

A focused pathway analysis of key metabolites. The fold changes of detected metabolites in Calvin cycle, sugar metabolism, C2 redirection pathway and squalene biosynthesis pathway during daytime between C2 redirection lines (pT5) and chloroplast functional FPS/SQS (G1) lines were highlighted by color according to the statistical significance calculated using Welsh's t-test. A decrease in sugar biosynthesis is observed, while malate and pyruvate levels are enriched, along with the engineered terpene bioproduct, squalene.

Figure 4.


O2
P-glycolate
CO2
RuBP
Glycolate
Rubose-5P
3P-glycerate
Glyoxylate
C2
Redirection
Pathway
Ribose-5P
Calvin
Cycle
1,3 bis P-glycerate
Malate
S-7P
G3P
Pyruvate
SBP
Erythrose-4P
1,6 bis-P-fructose
DXP
F6P
MEP
MEP
Pathway
Mannose-6P
Fructose
G-6P
HMBPP
Sugar
Metabolism
Mannose-1P
Sucose
Glucose
DMAPP
IPP
FPP
Squlene
Biosynthesis
Pathway
Increased (p<0.05)
Decreased (0.05<p<0.1)
No change (p>0.1)
Dcreased (p<0.05)
Squalene

Materials and Methods

Plant Materials

The transgenic FPS/SQS (FS) line overexpressing the FPS and SQS genes in this study was generated in previous studies [50, 51]. All vectors were transformed into the G1 line in this study containing chloroplast targeted FPPS and SQS coding sequences.

Expression Vector Construction

The *GDH* and *CAT* genes were codon-optimized and synthesized by Genscript. The *GO* gene (GenBank ID: NM_180254.2) was cloned from *Arabidopsis thaliana*. The pumpkin MS gene (GenBank ID: X56948.1) was amplified from pGreenII(hyg)-MS construct provided by Dr. Veronica Maurino (Heinrich Heine University). The DXPS gene was amplified from pDXS (codon-opt). The potato chloroplast transit peptide (tp) sequence (GenBank ID: EU294362) was amplified from pA215-036 provided by Dr. Christoph Peterhansel (Leibniz University). The cloning strategy for building the constructs was similar to a previous report[52]. All the genes were first fused with tp (transit peptide) sequence and inserted into a multi-cloning site flanked by corresponding promoters and Tnos terminators in pUC19. Then, the entire expression cassette of each gene, including the promoter, the coding sequences and the terminator, was amplified by corresponding forward primer and reverse primer in Table S1. An AvrII site was introduced into the forward primers, and the SpeI and AatII sites were introduced into reverse primers. In this way, the expression cassettes were flanked by AvrII-AatII, and sequentially cloned into the SpeI-AatII sites of pMHSuK. All of the genes and vectors were verified by sequencing.

Tobacco Transformation

The constructs were transformed into the *Agrobacterium tumefaciens* lab strain GV3101. Tobacco was transformed via the leaf-sac method[53]. A culture of *A. tumefaciens* was initiated from glycerol stock and grown overnight at 28°C with shaking (180 rpm) in liquid Luria-Bertani (LB) medium containing 50 mg/L kanamycin, to mid-log phase (OD_{600} = 0.5). The *A. tumefaciens* cells were collected by centrifugation for 10 min at 4,000 rpm and resuspended in liquid inoculation medium (containing MS salts, vitamins and 30 g/L sucrose). The *Agrobacterium* cell density was adjusted to an OD_{600} of 1.0 for inoculation. Excised leaves of 14-day-old G1 tobacco seedlings were used as the explant material for co-cultivation with *A. tumefaciens* GV3101 harboring the expression vectors. The excised explants were dipped into the *A. tumefaciens* culture in liquid inoculation medium for 20 min, blotted dry on sterile filter paper, and incubated in the dark at 25°C on agar-solidified MS medium with 2,4-D 2 mg/Ll. After 2 days of co-cultivation, the explant tissues were washed with sterilized distilled water three times and transferred to MS medium containing salts, vitamins, 6-Benzylaminopurine (BAP) 2 mg/L, sucrose 30 g/L, cefotaxime 500 mg/L, kanamycin 75 mg/L, and plant agar 8 g/L. Putative plants were observed emerging from the wound sites after 3-4 weeks. Selected plants were transferred to rooting media (MS basal salts with IAA 1 mg/L). The rooted plants were transferred to culture vessel for further growth.

Molecular Characterization

DNA was extracted from transgenic tobacco using the CTAB method[54]. Total RNA was extracted using the Quick-RNA Miniprep Kit (Zymo Research), and 2 μg of total RNA was reverse

transcribed with oligo-dT primers and SuperscriptII reverse transcriptase (Life Technology). Genomic DNA PCR and RT-PCR were carried out as previously described [55].

Squalene Quantification

Five hundred milligram of positive transgenic tobacco leaf tissue was collected using a 2 cm diameter cork borer. Each sample was ground with liquid nitrogen, and then extracted with 3 mL hexane. Cedrene was added at a final concentration of 9 ppm to serve as an internal control. The extracts were purified by passing through a 500 mg silica column in a glass pipette plugged with glass wool. Four ml of additional hexane was used to wash the column. The flow-through was analyzed by gas chromatography-mass spectrometry (GC-MS).GC-MS analysis was run on GCMS-QP2010SE (Shimadzu). One microliter of sample was injected into the GC-MS using an AOC-20i auto-sampler in 10:1 split mode (injector 280°C) onto a ZB-5MSi fused silica capillary column (30 m × 0.25 mm × 0.25 μm thickness). The initial oven temperature was 40°C, which was ramped to 120°C at 20°C/min, then ramped to 200°C at 6°C/min, then ramped to 260°C at 20°C/min, and finally ramped to 310°C for 3 min at 5°C/min. Helium was used as the carrier gas. The ion source was set to 230°C and the interface was 280°C. Squalene quantification was performed using selected ions. Peak identification of the compound was performed using direct comparison of the sample mass chromatogram with those of commercially available standard compounds. The quantitative calculations of squalene concentration were based on the peak area ratios relative to those of the standard.

C_{14} Labeled Glycolate Feeding Assay

About 10 leaf discs (~50 mg) were collected using a 2 cm diameter cork borer. Each disc was placed onto 100 μL water containing 0.5 μCi [1,2-^{14}C]glycolate on a parafilm-lined petri dish.

Approximately 500,000 to 1,000,000 cpm per 100 µL were used. Leaf discs were placed on a light bank shelf and collected at 0, 12 and 24 hours. When the materials were collected, both sides of leaf discs were washed with 3 mL water, blotted dry, quickly frozen in liquid nitrogen, and stored at -20°C.After that, the samples were powdered in liquid nitrogen and extracted with 3 mL of hexane containing 200 ng of squalene to serve as a carrier. The samples were shaken for 2 hours at room temperature, and then the organic extract was transferred to a clean glass vial and concentrated to dryness under nitrogen gas. The samples were resuspended in 100 µL hexane, and 20 µL of sample was loaded separated on silica TLC plates (60 Å silica gel). The TLC plates were stained with iodine vapor to locate squalene zone, and the silica in the squalene zone was scraped from the plate and counted in a scintillation counter to determine the radioactivity.

Metabolomics

Four biological repeat of samples were collected from the parental FS line (G1) and pRD4A (photorespiratory bypass line) T_1 plants in the morning (10:00 am) and night (10:00 pm). The procedures for metabolomic profiling have been described before [56]. Briefly, the samples were prepared using the automated MicroLab STAR® system (Hamilton). A recovery standard was added prior to the first step in the extraction process for QC purposes. The extract was analyzed by three independent platforms: one analysis by UPLC-MS/MS with positive ion mode electrospray ionization, one analysis by UPLC-MS/MS with negative ion mode electrospray ionization, and one analysis by GC-MS.The LC-MS of the platform was based on a Waters ACQUITY ultra-performance liquid chromatography (UPLC) and a Thermo-Finnigan LTQ mass spectrometer operated at nominal mass resolution, which consisted of an electrospray ionization (ESI) source and linear ion-trap (LIT) mass analyzer. The sample extract was dried then reconstituted in acidic or basic LC-compatible solvents, each of which contained 12 or more

injection standards at fixed concentrations. One aliquot was analyzed using acidic positive ion-optimized conditions and the other using basic negative ion-optimized conditions in two independent injections using separate dedicated columns (Waters UPLC BEH C18-2.1×100 mm, 1.7 μm). Extracts reconstituted in acidic conditions were gradient eluted using water and methanol containing 0.1% formic acid, while the basic extracts, which also used water/methanol, contained 6.5 mM ammonium bicarbonate. The MS analysis alternated between MS and data-dependent MS/MS scans using dynamic exclusion and the scan range was 80-1000 m/z. Raw data files were archived and extracted as described below.The samples destined for analysis by GC-MS were dried under vacuum for a minimum of 18 h prior to being derivatized under dried nitrogen using bistrimethyl-silyltrifluoroacetamide. Derivatized samples were separated on a 5% diphenyl / 95% dimethyl polysiloxane fused silica column (20 m x 0.18 mm ID; 0.18 μm film thickness) with helium as carrier gas and a temperature ramp from 64°C to 340°C in a 17.5 min period. Samples were analyzed on a Thermo-Finnigan Trace DSQ fast-scanning single-quadrupole mass spectrometer using electron impact ionization (EI) and operated at unit mass resolving power. The scan range was 50–750 m/z. Raw data files are archived and extracted as described below. Metabolites were identified by automated comparison of the ion features in the experimental samples to a reference library of chemical standard entries that included the retention time/index (RI), mass to charge ratio (m/z), and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Peaks were quantified using area-under-the-curve. The raw data were normalized against the total ion count, scaled against the median of a metabolite and curated. Statistical analysis of the data was performed using JMP (SAS, http://www.jmp.com), and R (http://cran.r-project.org/). A log transformation was applied to the observed relative

concentrations for each biochemical. Welch's two-sample t tests were used to determine whether each metabolite was significantly increased or decreased in abundance.

III. ENVIRONMENTAL EVALUATION OF C2 PHOTORESPIRATORY BYPASS

Introduction

To better understand the environmental impact on efficiency of C2 photorespiratory redirected carbon flux, a controlled environment experiment was conducted. The environments varied in temperature, humidity, and irradiance to produce a low photorespiration environment as a control, and an elevated photorespiration chamber to evaluate the C2 photorespiration bypass efficacy. Both G1 (ctpFPPS+ctpSQS) and photorespiratory bypass plants were evaluated from juvenile stage to vegetative maturity (pre-flowering) in both environments. Quantification of the engineered bioproduct squalene, photosynthetic rates, and a metabolomics analysis were performed to determine variation between genotypes and the impact the growth condition has on squalene yield as well as insight to genetic and environmental interactions. Photorespiration byproducts increase under increasing heat and irradiance[57], allowing a larger pool of carbon for rechanneling into terpene. However elevated photorespiration rates also reduce photosynthesis[58], reducing the overall fixed carbon and thus the allocation of carbon within the plant's primary, or in our case, engineered secondary metabolism for terpene biosynthesis. A plant's phenotype, as any organism's phenotype, is result of both the organism's genotype and environment[59]. One thing that makes plants unique to many other forms of life is their sessile nature. This means avoidance of many biotic and abiotic stress must be done internally through ramped up defenses and oxidative stress mitigation strategies[60]. A major problem with a sessile autotroph is regulation of incoming oxidation energy and the fixation of new carbon for primary metabolism. Excess irradiance brings more photons than the photosystem can process, leading to photoinhibition, halting photosynthesis temporarily[58]. Excess oxidation energy also decreases

the reduction potential within the stroma, decreasing intracellular CO_2 concentrations and increasing the oxygenic reactions of RuBisCo, known as photorespiration[57]. The phenotype equation is used in plant breeding to estimate the amount of gain in a particular phenotype realized from genetic components (additive genetics) or environmental components (growth condition). This is done by comparing genotypes in varied environments or growth conditions[59]. After testing the C2 photorespiratory bypass plants in the greenhouse with modest gains over the engineered terpene sink only lines, it was hypothesized that the higher photosynthesis rates in full sun light and increased photorespiration would lead to both an increase in squalene as well as an increasing gain observed over the G1 control lines. The full sun environment did increase squalene yields overall as expected, however gain from C2 photorespiratory bypass lines over terpene biosynthesis sink only plants was mitigated. Furthermore, variability between lines was high. The condition of the material throughout the growth season of course changed, as older tobacco begins to produce a leathery quality nearing maturity and often loses some pigmentation nearing harvesting and post-suckering. Interestingly, whether or not overall variability was high between lines, or more or less gain of the C2 photorespiratory bypass design over other lines, the bypass plants always produced a higher variation within their own group. Initially this was thought to be unstable lines of a transgenic event, however this variation was not observed in other lines of similar generations. After multiple generations, the phenotype variation still exists. More importantly, based on photorespiration mechanisms, the ratio between oxidation and fixation always depend heavily on the environment. This meant yield prediction for these lines was more difficult than other lines as well as being more strongly acted upon by the environmental factors of their temporal growth conditions than the (dominate and additive) transgenic cassettes separating them from terpene biosynthesis sink plants.

The C2 photorespiratory bypass was constructed to re-route carbon from p-glycolate to the beginning of the MEP pathway. The modeling and labeled carbon feeding assays performed by Hong Ma show an increase in squalene content as photorespiration increases. This model does not however suggest an optimal photorespiration rate nor ratio to photosynthesis. The intrinsic connection between the MEP and photosynthesis have been proven and are well understood, requiring NADPH and ATP from light reactions and G3P from the Calvin Benson cycle to produce terpene precursors DMAPP and IPP[47]. The constitutive engineered terpene biosynthesis sink also requires a heavy demand for carbon. The photorespiratory bypass seeks to capitalize on the inefficiency of photorespiratory reactions by shortening the path of detoxification of p-glycolate and directing photorespiratory carbon to the MEP pathway. Modeling suggests that an increase in photorespiration would result in an increase of carbon being re-routed and thus a gain in bioproduct yield in a linear correlation. Experimentally we see an increase in the photorespiratory bypass lines over lines only containing the downstream terpene biosynthesis sink, but attribution of this yield increase to photorespiration alone or rerouted carbon has been difficult to prove.

The overall scheme of the photorespiratory bypass routes p-glycolate toward DXP, through which a two carbon molecule is re-routed to a five carbon compound. In the process however, at the synthesis of pyruvate from malate and in the condensation of G3P and pyruvate, two CO_2 are evolved. Net carbon gain into the MEP therefore is theoretically zero. In this scheme the bypass functions by removing the product of RuBisCO's oxygenase reactions and yields two molecules of CO_2, enriching the fixation environment in which RuBisCO is catalyzing reactions. The photorespiratory bypass functions as or at least is analogous to a carbon concentrating mechanism within the stroma. Yield gains made from the photorespiratory bypass may be attributed to both rerouting carbon, largely sugars from primary metabolism due to the strong downstream sink with

help by malate synthase within the bypass design itself, as well as increasing intracellular or more accurately inner chloroplast CO_2 concentrations.

From previous study we are aware that the terpene sink only lines downregulate sugar biosynthesis as carbon flux is pulled to the strong terpene synthesis sink competing with primary metabolism. We also observed this trend in the C2 photorespiratory bypass, to greater effect as G3P is siphoned off into the condensation reaction with pyruvate to produce DXP for the MEP pathway. One could theorize that with increasing photosynthesis rate, carbon economy within the plant would increase and rates of photorespiration would increase. However, for any given RuBisCO reaction, photorespiration and photosynthesis are antagonistic and mutually exclusive. The working conclusion for this remains, as photosynthesis rate increases: (1) Biosynthesis of terpene bioproduct in both lines increases, (2) Gain realized from photorespiration in the bypass line decreases as it is diminishing scaled to photosynthesis impact on net carbon gain. Inversely, one may expect an increase in photorespiration would lead to an increased realized gain from the photorespiratory bypass over the terpene sink only lines when the ratio of photorespiration to photosynthesis is high, but overall photosynthesis rates are low. Also, due to the carbon concentrating like mechanism of the bypass design evolving CO_2, bypass plants may also realize a gain under low photosynthetic potentials by improving stromal CO_2 concentrations thus increasing fixation efficiency[61]. In order to test these theories of environmental control on carbon flux to terpene bioproduct, we examined squalene content, photosynthetic rate, and global metabolic changes between the C2 photorespiratory bypass design and plants containing just the terpene biosynthesis sink.

Materials & Methods

Plant Materials

Plants used in this study were germinated onto MS +Kan_{100} (Kanamycin 100mg*L^{-1}) selection media for the bypass T-DNA insert, although the seed had been derived from homozygous transgenic lines and previously confirmed. The C2 photorespiratory bypass seed came from T_4 generation pT5 lines 38 and 40, supplied by Hong Ma. Terpene sink only seed, from T_6 homozygous G1, the parent line of the pT5 transformation events. Seeds for G1 were germinated in MS +Hyg_{75} (Hygromycin 75mg*L^{-1}). The healthiest 12 seedlings were selected to be transferred to soil in 2" pots under a humidity dome, about 2 weeks post germination. After hardening off and opening their 3rd set of leaves, plants were moved to the greenhouse for growth and stiffening of their stems for an additional 2 weeks. After this first month of growth, squalene quantification was conducted to group the median producing plants for each genotype to use in the controlled environment. A baseline sample was taken for quantification after 24hrs. Plants were allowed to then grow, adapt, and age over the next month. At two months old, and one month in the controlled environments squalene quantification was performed every other week, at vegetative maturity and before initiating floral development.

Controlled Environment

Two different cultivation environments were simulated using environmentally controlled growth chambers. One with low irradiance, cool temperatures, and moderate humidity mimicking environmental conditions observed in US tobacco cultivation early spring. The other with increased irradiance and high temperatures to increase photorespiration along with reduced humidity to restrict stomatal gas exchange, increasing internal O_2:CO_2 concentrations[62]. The

increased photorespiration environment (chamber B13) had a mean temperature of 34.5±0.18°C, relative humidity of 31±1.59%, and 17,000 lux (lumen*m^{-2}) at canopy height (1m from floor) with a 16hr light: 8hr dark cycle. The low photorespiration environment (chamber B10) had a mean temperature of 21±0.07°C with relative humidity at 61±0.90% and canopy light intensity during light hours of 4,300 lux with a 16:8hr photoperiod. B10 light intensity was roughly half of "full daylight", and 4-fold of an overcast day, whereas B13 illumination exceeds lux of full daylight. "Full sunlight", or direct solar irradiance on a cloudless day, differs still almost 10-fold of what is achievable within the growth chambers, and is typically estimated at sea level to be closer to 100,000 lux. For comparison of light conditions, lumens and proportional lux, reference (**Table 1**). The temperature for B13 was raised just above mid-summer average daily high temperatures in US tobacco cultivation areas by a few degrees Celsius, while B10 mimics an early spring average around April after transplants typically go to the field for cultivation. The relative humidity condition of B10 also works to mimic outdoor growth conditions, where in the spring humidity is around 70-80% in the morning, and 40-50% in the afternoon, with the average typically around 60% (*61% for Greensboro, NC 2018*). Humidity for environment B13 was reduced to 30% in efforts to reduce gas exchange and increase photorespiration rates. This allowed us to reduce the exchange of gas and thus decrease intercellular CO_2, in the effort to stimulate photorespiratory products within limited light intensities achievable within the growth chambers[62].

Table 1. Relative lux ranges and conversion chart

Approximate lux output correlating to different environmental illumination conditions provided by the National Optical Astronomy Observatory. 1 lux =1 lumen*m-2, allowing for light energy to be measured at canopy height instead of by light source output.

Condition	**Illumination**	
	(ftcd)	***(lux)***
Sunlight	10,000	107,527
Full Daylight	1,000	10,752
Overcast Day	100	1,075
Very Dark Day	10	107
Twilight	1	10.8
Deep Twilight	.1	1.08
Full Moon	.01	.108
Quarter Moon	.001	.0108
Starlight	.0001	.0011
Overcast Night	.00001	.0001

- *1 lux = 1 lumen / sq meter = 0.0001 phot = 0.0929 foot candle (ftcd, fcd)*
- *1 phot = 1 lumen / sq centimeter = 10000 lumens / sq meter = 10000 lux*
- *1 foot candle (ftcd, fcd) = 1 lumen / sq ft = 10.752 lux*

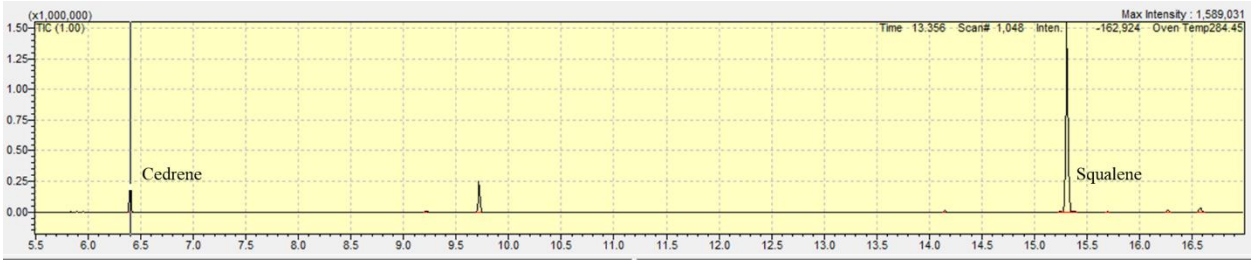
In each environment, four biological replicates of photorespiration bypass plants and two biological replicates of G1 (ctpFPPS+ctpSQS) without the bypass, totaling six in each chamber, eight total pT5 individuals, and 4 total G1 individuals.

Squalene Quantification

From these plants, biweekly tissue samples were excised for squalene extraction, with two separate replicated extractions per each plant. This gives us four biological replicates for each environment for the bypass lines, two each for the terpene sink only controls, and two technical replicates for each biological replicate. As within prior study, the material harvested for squalene analysis came from the first mature leaf from the top, as younger leaves are strong sink organs which consume sugars for growth, and older leaves have drop off in photosynthetic capacity as well as partial shading. Both 0.5g fresh leaf samples (to compare with previous data sets) and 0.1g lyophilized leaf (reduced variation from water content) tissue were macerated to powder with a glass pestle after using liquid N_2 to flash freeze samples in sealed 50ml falcon tubes. After maceration, samples were extracted using HPLC grade (98.5%) n-hexanes by shaking powdered leaf material and 3ml n-hexanes in an orbital shaker for 2.5hrs. Cedrene was added to the hexane extraction prior to shaking at a final concentration of 15 ppm to serve as a spiked internal standard for quantification from MS peak areas. Cedrene being a terpene gives it similar extraction and degradation properties to squalene, but being a sesquiterpene it has half the mass and therefore does not overlap peaks in the GC/MS chromatograph (**Figure 5**). This allows quantification of squalene, and factoring for degradation by heat, hydrogenation, or oxidation during the extraction process. The extracts were then filtered by passing through a 3.4cm silica (size 60, 40-63µm diameter) column in a glass pasteur pipette plugged with glass wool at the bottom and a thin layer of sodium sulfate ($NaSO_4$) on top.

Figure 5. Squalene Quantifying Chromatogram

Chromatogram of squalene quantification using Cedrene as an extraction standard. Cedrene added to the hexane extraction to a final concentration of 15ppm is observed at 6.400min, and squalene at 15.310min. The given quantification for Cedrene allows quantification of the area under the curve for the squalene peak, identified by its ion masses.



Half (1.5ml) of the extraction was pipetted into the top of and passed through the silica chromatography column. Then the column was washed with hexane until the final eluted collection volume reached 6ml, allowing for all squalene to have passed through into the 10ml glass vial. A visual marker of this completion is the yellow/orange band of beta-carotene, a tetraterpenoid, reaching the glass wool. From here the 6ml final collection volume was adjusted using N_2 gas over the open vial head space, giving consistent volume and a non-oxidative environment to the filtered non-polar eluent. From this, 1ml was put into GC/MS vials and each sample was run on the GC/MS. GC/MS analysis was run on GCMS-QP2010SE (Shimadzu). One microliter of sample was injected into the GC/MS using an AOC-20i auto-sampler in 30:1 split mode (injector 220°C) onto a ZB-5MSi fused silica capillary column (30m × 0.25mm × 0.25µm thickness). The initial oven temperature was initially set to 40°C with a 30sec hold, then ramped up to 165°C at a rate of 30°C/min, then to 185 at a rate of 5°C/min, then to 280°C at a rate of 25°C/min, and finally up to 300°C at a rate of 5°C/min with a 2min hold. Helium was used as the carrier gas. The ion source was set to 230°C and the interface was 280°C. Squalene quantification was performed using selected ions. Peak identification of the compound was performed using direct comparison of the sample mass chromatogram with those of commercially available standard compounds. The quantitative calculations of squalene concentration were based on the peak area ratios relative to those of the standard (15ppm cedrene). To ensure the precision of the measurements, one sample from each extraction batch was run as a technical replicate (*from same 6ml vial and ran on GCMS during each processing batch*) to ensure quality of measurements. The GC/MS squalene yield in ppm, relative to the Cedrene 15ppm, was then converted to µg/g fresh weight (FW) and mg/g dry weight (DW) of squalene. Many calculations were kept in ug/g FW/DW although yields were in the mg/g range to preserve accuracy and reduce any rounding errors in subsequent data analysis.

This standardized comparison of yield data between lines, events, and constructs, is helpful to calculate bioproduct to biomass ratios. The latter is needed for project feasibility estimates and to predict pilot study yields. Mass to mass ratios also allow us to express the bioproduct yield as a percent mass of the biomass, making estimation on larger scale production schemes simpler.

Photosynthesis Measurement

Photosynthetic activity was measured for both genotypes in both treatments, using a LI-COR® LI-6400 portable photosynthesis system from LI-COR® Biosciences. The LI-6400 yields measurements of photosynthesis and transpiration based on the differences in CO_2 and H_2O in an air stream that is flowing through the leaf cuvette. After 2 weeks in the controlled environments, photosynthesis measurements were taken on the first mature leaf from the top for each plant. Measurements were completed five hours after the light cycle had begun, when photosynthesis reaches its peak efficiency and well before midday depression of photosynthetic rates. The median of 3 measurements was taken as the read for each individual, and biological replications equaled the number of individuals. Mean photosynthesis rate for each genotype in both environmental treatments was then calculated for comparison to modeling and yield data.

Metabolome Methods

Leaf tissue samples collected for the third squalene quantification was also used to perform metabolomics. Samples were pooled by genotype and environmental condition, yielding four individual pooled samples: B10-pT5, B10-G1, B13-pT5, and B13-G1. These pooled samples were sent to Metabolon® for principal component and metabolomics analysis. Samples were prepared using the automated MicroLab STAR® system from Hamilton Company. Several recovery standards were added prior to the first step in the extraction process for QC purposes. To remove

protein, dissociate small molecules bound to protein or trapped in the precipitated protein matrix, and to recover chemically diverse metabolites, proteins were precipitated with methanol under vigorous shaking for 2 min (Glen Mills GenoGrinder 2000) followed by centrifugation. The resulting extract was divided into five fractions: two for analysis by two separate reverse phase (RP)/UPLC-MS/MS methods with positive ion mode electrospray ionization (ESI), one for analysis by RP/UPLC-MS/MS with negative ion mode ESI, one for analysis by HILIC/UPLC-MS/MS with negative ion mode ESI, and one sample was reserved for backup. Samples were placed briefly on a TurboVap® (Zymark) to remove the organic solvent. The sample extracts were stored overnight under nitrogen before preparation for analysis. Raw data was extracted, peak-identified and QC processed using Metabolon's hardware and software. These systems are built on a web-service platform utilizing Microsoft's .NET technologies, which run on high-performance application servers and fiber-channel storage arrays in clusters to provide active failover and load-balancing. Compounds were identified by comparison to library entries of purified standards or recurrent unknown entities. Metabolon® maintains a library based on authenticated standards that contains the retention time/index (RI), mass to charge ratio (m/z), and chromatographic data (including MS/MS spectral data) on all molecules present in the library. Furthermore, biochemical identifications are based on three criteria: retention index within a narrow RI window of the proposed identification, accurate mass match to the library +/- 10 ppm, and the MS/MS forward and reverse scores between the experimental data and authentic standards. The MS/MS scores are based on a comparison of the ions present in the experimental spectrum to the ions present in the library spectrum. While there may be similarities between these molecules based on one of these factors, the use of all three data points can be utilized to distinguish and differentiate biochemicals. More than 3300 commercially available purified standard compounds

have been acquired and registered into Metabolon's software for analysis on all platforms for determination of their analytical characteristics. Additional mass spectral entries have been created for structurally unnamed biochemicals, which have been identified by virtue of their recurrent nature (both chromatographic and mass spectral). These compounds have the potential to be identified by future acquisition of a matching purified standard or by classical structural analysis. Identified compounds and their up and down regulated fold changes and variances are then reported for interpretation along with other data collection methods.

Results and Discussion

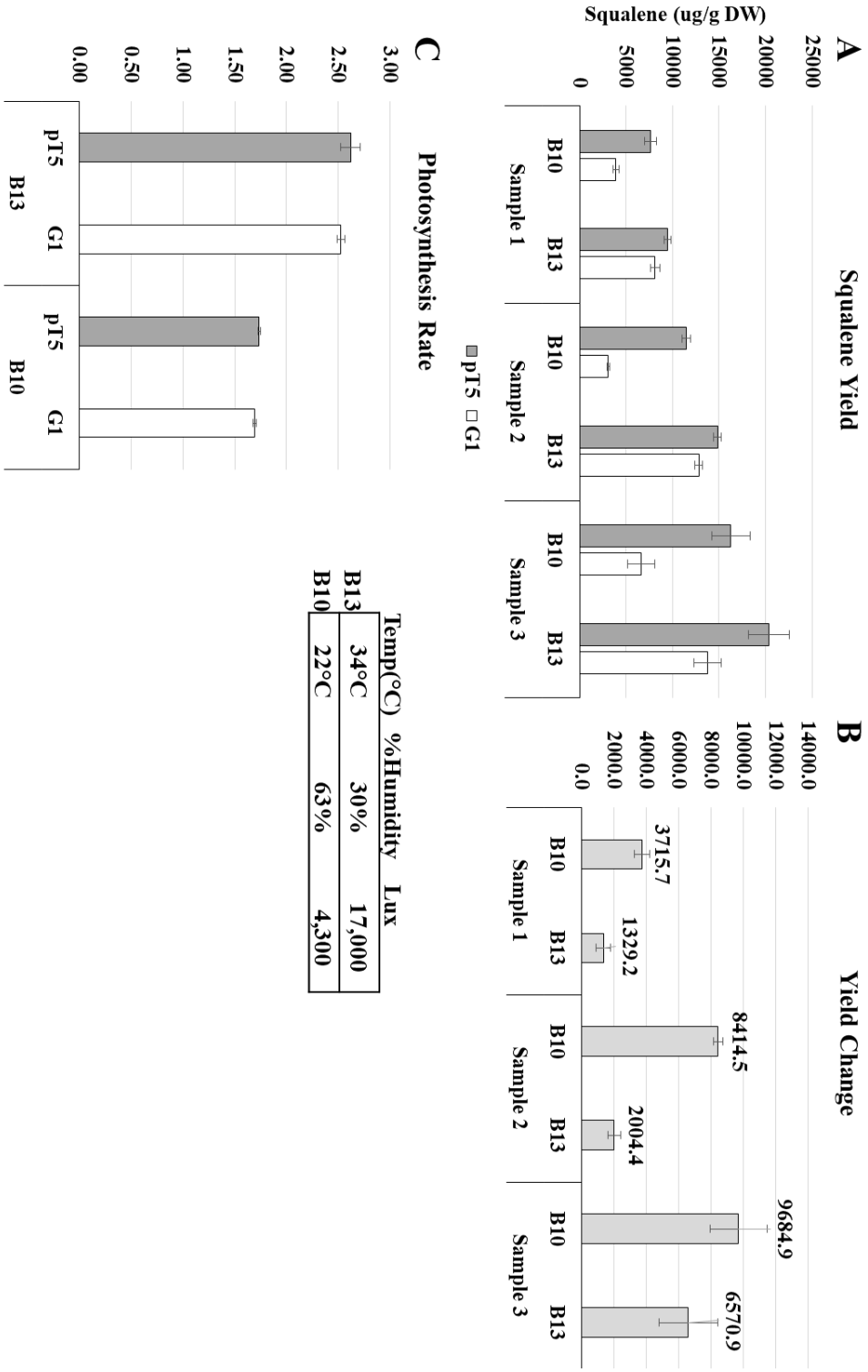
Squalene Quantification

Quantification of squalene bioproduct showed the modeling of the C2 photorespiratory bypass pathway was correct in predicting increasing carbon flux to terpene biosynthesis with increasing photorespiration, but the model fails to quantify that gain relative to net carbon gain from photosynthesis (**Figure 6A**). The relative accumulation of the engineered bioproduct increases of course with increased photosynthesis, but the ratio of photosynthesis to photorespiration greatly impacts the gain that can be realized from the bypass pathway. Under increased irradiation such as in condition B13, photosynthesis was increased leading to more carbon allocated to secondary and even tertiary metabolites such as the terpene sink, making the small amount of rerouted carbon from the bypass contribute much less to overall carbon allocation than under more limited photosynthesis environment B10 (**Figure 6B**).

Figure 6. Squalene Quantification, Difference, and Photosynthesis Rates for pT5 and G1 in B10 and B13

Quantification of (**A**) squalene yield and (**B**) yield gain of pT5 (C2 photorespiratory bypass into G1) over G1 (ctpFPPS+ctpSQS) while grown in controlled environment growth chambers, B10 a low photorespiratory treatment and B13 a high photorespiratory treatment. Samples every two weeks shows increase in yield attributed to growth and maturity. The C2 redirection increases squalene yield, but with reduced yield gains in condition B13. (**C**) Photosynthesis rates of both genotypes in both treatment groups. No significant difference was observed between genotypes in the same treatment group.

Figure 6.


A
Squalene Yield
Squalene (ug/g DW)
25000
20000
15000
10000
5000
0
B10
B13
Sample 1
Sample 2
Sample 3
pT5
G1
B
Yield Change
14000.0
12000.0
10000.0
8000.0
6000.0
4000.0
2000.0
0.0
3715.7
1329.2
8414.5
2004.4
9684.9
6570.9
C
Photosynthesis Rate
3.00
2.50
2.00
1.50
1.00
0.50
0.00
pT5
G1
B13
B10
Temp(°C) %Humidity Lux
B13 34°C 30% 17,000
B10 22°C 63% 4,300

The C2 photorespiratory bypass plants outperformed the terpene biosynthesis sink only plants in both conditions and at each time point. As expected, and observed throughout this and previous work, squalene yield varied by age and maturity of the plant and leaf material sampled. The young leaves, while having a strong carbon sink strength and dense with chloroplasts, typically take time to accumulate the engineered terpene bioproduct. At maturity, still photosynthetically active leaves have their highest titer and most stable yield. We also know squalene diffuses out of the chloroplast membrane, leading to a reduction during dark hours and a peak yield late into the light cycle[46]. The gain observed in the same treatment and group over the three sampling times is attributed to growth and maturity of the plant and leaf material, pre-flowering. The gain realized over the terpene biosynthesis sink only plants by the C2 photorespiratory plants observed in the different treatments suggests that while the C2 bypass pathway does work, there is a limited flux when photorespiration increases. One issue is the still active *plgg1* transporter, allowing traditional schemes of photorespiration to occur[63], and reducing the effectiveness of the bypass under increasing photorespiration[35]. More interesting than the diminished gain observed in treatment B13, is the additive effect on yield observed in the low irradiance environment. Individual components within the C2 photorespiratory bypass design are likely contributing to a more reduced environment for fixation and producing CO_2 within the stroma, leading to the bypass lines having higher intracellular CO_2 concentration.

Photosynthesis Quantification

Photosynthetic output was quantified in order to determine if there were differences in the CO_2 uptake and stomatal conductance between the two genotypes in each treatment. As expected the plants in the higher irradiance had increased photosynthesis rates, albeit 10-30% of what is observed in the field under full sun light. The comparison is shown in **Figure 6C**, where as

predicted, B13 treatment shows increase over B10. Not observed was a statistical difference between genotypes within the same treatment. Modeling predicts no net change in photosynthetic rate as does our experimental observations, however other groups have observed statistically significant variation in the C2 photorespiratory bypass design and synthetic glycolate metabolism, as well as large significant variation when coupled with a *plgg1* knock down[35]. The change in photosynthesis by such a design hinges on reducing or stopping traditional energy dissipating models of photorespiration producing a shorter functional route for p-glycolate to be reintroduced to primary metabolic functions.

Metabolome Results

One way to judge the magnitude of metabolic effects caused by specific variables is to simply summarize the number of compounds significantly changed ($p \leq 0.05$) in each pair-wise comparison, and the direction of those changes. However, pooled samples for both genotype and condition produced artificially low intragroup variation due to being technical replicates. This lead to many compounds meeting the statistical cut-off. Still, the larger absolute differences between compound levels caused by the environmental component resulted in many more compounds being statistically significant in the environmentally impacted group of compounds. Because of this increased discovery rate for compounds within these technical replications we categorize the various correlations in terms of fold-change, assuming the largest differences between single samples have the greatest chance of representing actual biological effects, instead of using a p-value cut-off to characterize the results. This focuses our efforts on the largest impacted and most altered metabolites within comparisons, either by genotype or where most variation occurred within the environment treatment variable. **Table 2** lists the environmental treatment related fold changes in both the pT5 (photorespiratory bypass) and G1 (terpene sink only) background plants,

such that the effects of environment B13 are compared to that of B10 (B13/B10). Included are all those changes which were >3-fold in either direction in both genotypes. Several compounds increased by 10-fold or more, some reaching 100-fold or more. The most notable observations were that the higher temperature and adequate light led to much greater carbon accumulation, active membrane remodeling, and higher amino acid catabolism, all effects expected in a more actively growing tissue.

Sucrose represented the greatest difference correlated to the environmental difference, being >100-fold higher for pT5 in environment B13 relative to B10, and >40-fold higher for G1 in B13 relative to B10 (**Figure 7**). Pyruvate, was elevated multiple fold in the B13 condition for both genotypes along with storage pathway metabolites derived from sucrose, such as 1-kestose and galactinol. Glucose and fructose, which can derive from invertase activity, were slightly elevated in the B10 and B13 samples, but fell below the >3 fold threshold. Similar behavior was observed for sugars connected to the pentose phosphate pathways (**Table 3**), including; xylulose, xylose, ribose, ribulose, and sedoheptulose. Rearrangement and phosphorylation of C5 sugars make fold changes in either direction of a single C5 species less likely. Other metabolic impacts included decreased IAA and its derivatives by nearly 5 fold in B10 compared to B13. IAA metabolism also exhibited a slight genotypic effect, being slightly elevated in the photorespiration bypass genotype, although the majority of variation exists in the environmental treatment. Another clear effect of the higher light and temperature was higher levels of ascorbate. The precursor species, gulonate, was not significantly different in the two conditions, suggesting that levels of ascorbate in the system may be regulated by reduced catabolism.

Table 2. Environmental Impact on Metabolomes of pT5 and G1

List of the environmental treatment related fold changes in both the pT5 (photorespiratory bypass) and G1 (terpene sink only) background plants, such that the effects of environment B13 are compared to that of B10 (B13/B10). Confidence interval of significance produced by Welch's t-test. Increased (red) and decreased (green) compounds less than 3-fold change in either direction may be significant statistically, but not metabolically as they are largely small changes in growth rate phenotype.

Table 2.

Super Pathway	Biochemical Name	B13 / B10	B13_G1 / B10_G1
Amino acid	phenylpyruvate	13.41	10.5
	4-hydroxyphenylpyruvate	4.19	4.4
	octopamine	12.92	4.74
	pipecolate	7.46	3.22
	1-ribosyl-imidazoleacetate*	14.16	7.91
	N-acetylarginine	7.12	4.62
	3-methyl-2-oxovalerate	8.62	4.23
	gamma-glutamylglycine	11	3.52
Carbohydrate	pyruvate	8.58	3.04
	myo-inositol	7.53	4.34
	galactinol	6.17	3.72
	sucrose	106.17	44.25
Lipids	linoleate (18:2n6)	3.34	3
	linolenate [alpha or gamma; (18:3n3 or 6)]	33.69	5.13
	1-palmitoleoyl-GPA (16:1)*	5.5	5.83
	1-linoleoyl-GPA (18:2)*	3.52	3.95
	1-linoleoylglycerol (18:2)	6.94	4.58
	1-linolenoylglycerol (18:3)	96.25	10.48
	2-linoleoylglycerol (18:2)	24.77	13.68
	1-palmitoyl-2-linolenoyl-galactosylglycerol (16:0/18:3)*	3.95	5.06
	1-linoleoyl-2-linolenoyl-galactosylglycerol (18:2/18:3)*	3.08	4.4
Cofactors, Prosthetic Groups, Electron Carriers	nicotinamide riboside	4.64	8.74
	dehydroascorbate	45.34	9.1
	pheophorbide A	3.08	4.42
Nucleotide	inosine	77.24	70.49
	N6-succinyladenosine	5.8	4.08
	orotate	9.38	5.13
	3-ureidopropionate	4.19	6.87
Peptide	phenylalanylalanine	9.12	8.62
	valylleucine	4.34	3.83
Secondary metabolism	rutin	5.46	6.9
Xenobiotics	phenylglucopyranoside	10.35	5.85

Super Pathway	Biochemical Name	B13 / B10	B13_G1 / B10_G1
Amino acid	serine	0.28	0.33
	asparagine	0.21	0.31
	homocitrate	0.28	0.17
	histidine	0.24	0.24
	N-acetylhistidine	0.31	0.24
	N-monomethylarginine	0.23	0.17
	3-hydroxy-2-ethylpropionate	0.17	0.11
	3-methylglutaconate	0.15	0.1
	gamma-glutamylhistidine	0.14	0.06
	gamma-glutamylthreonine	0.24	0.33
Carbohydrate	isocitrate	0.11	0.25
	N-acetyl-glucosamine 1-phosphate	0.28	0.29
Lipids	heptanoate (7:0)	0.11	0.21
	3-hydroxydecanoate	0.16	0.24
	azelate (C9-DC)	0.13	0.3
	1,2-dilinolenoyl-GPC (18:3/18:3)*	0.25	0.28
Nucleotide	2'-deoxyguanosine	0.11	0.25
	guanine	0.08	0.18
	5-methyl-2'-deoxycytidine	0.12	0.24
	thymine	0.18	0.26
Hormone metabolism	indoleacetate	0.2	0.12
Secondary metabolism	coumaroylquinate (2)	0.21	0.12
	coumaroylquinate (4)	0.23	0.1
	coumaroylquinate (5)	0.24	0.14
	coumaroylquinate (3)	0.17	0.12

0.55
0.76
1.71
1.32
1.20

Green: indicates significant difference ($p \leq 0.05$) between the groups shown, metabolite ratio of < 1.00

Light Green: narrowly missed statistical cutoff for significance $0.05<p<0.10$, metabolite ratio of < 1.00

Red: indicates significant difference ($p \leq 0.05$) between the groups shown; metabolite ratio of ≥ 1.00

Light Red: narrowly missed statistical cutoff for significance $0.05<p<0.10$, metabolite ratio of ≥ 1.00

Non-colored text and cell: mean values are not significantly different for that comparison

Figure 7. Box and whisker plots for selected key metabolites

Box and whisker plots for selected key metabolites. The largest fold change was observed in sucrose content. Glucose and fructose were also strongly influenced by environmental treatment. Sugar biosynthesis was decreased in B10 under restricted photosynthesis, and increased in B13 with higher photosynthesis. Pyruvate, malate, and glycerate all exhibited small genotypic differences as well as environmental differences, and are indicative of the genotypic contribution to metabolic flux changes from the C2 photorespiratory bypass engineering. A genotype and environment interaction is observed in pyruvate fold change, where in B10 consumption of pyruvate appears exacerbated possibly by DXPS.

Figure 7.

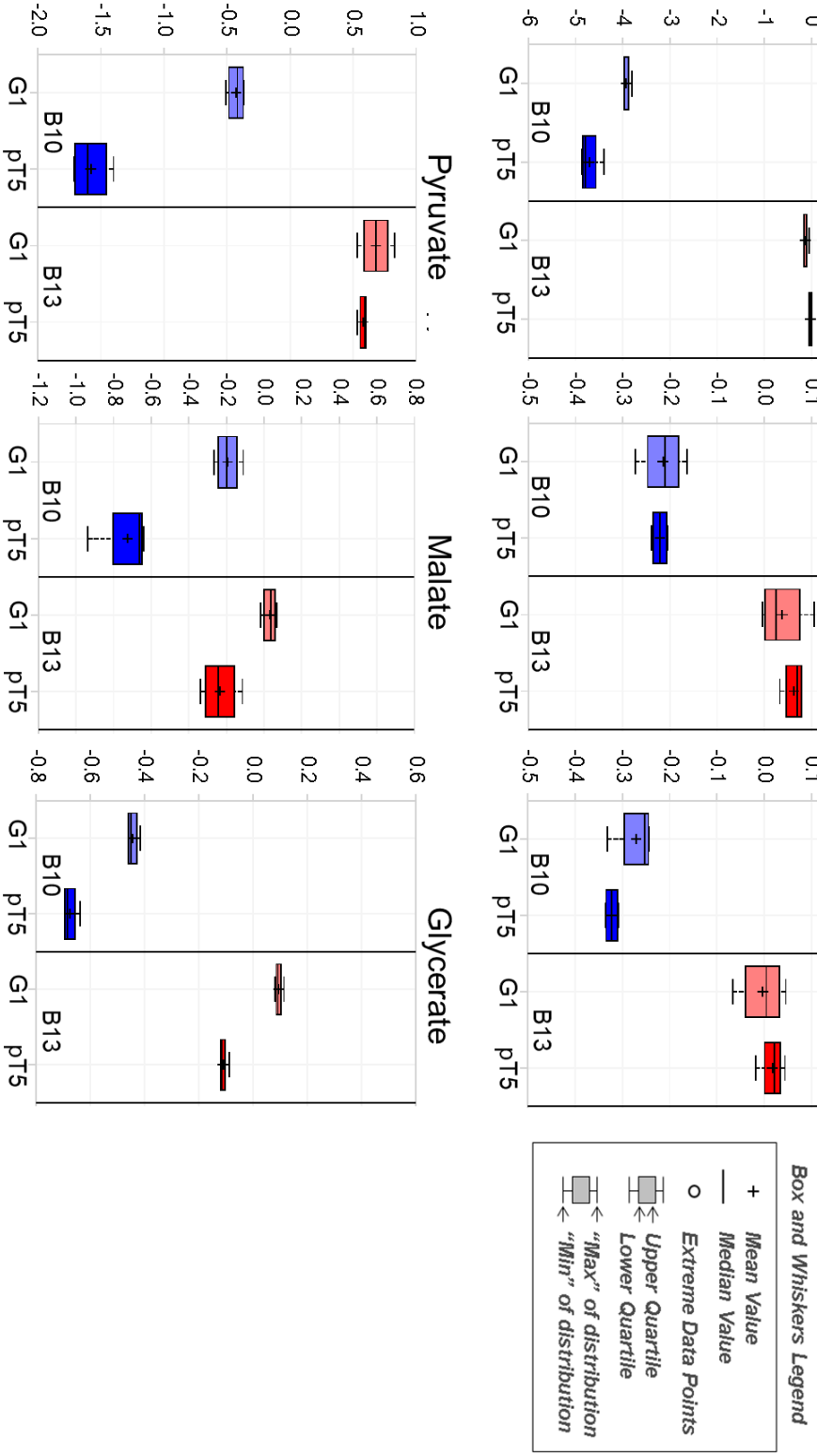

Sucrose
Glucose
Fructose
Pyruvate
Malate
Glycerate
G1
pT5
B10
B13
Box and Whiskers Legend
Mean Value
Median Value
Extreme Data Points
Upper Quartile
Lower Quartile
"Max" of distribution
"Min" of distribution

Table 3. Fold change of key metabolites in carbon flux scheme

Fold of change up or down along with Welch's t-test confidence intervals for key metabolites to the C2 photorespiratory bypass. Sucrose, which has the largest observed fold change, Glucose, and Fructose, are primary metabolism carbon sources. Ribulose/Xylulose, Xylose, Ribose, and Sedoheptulose are sugars which when phosphorylated are involved in the Calvin Benson cycle. Malate and Pyruvate are organic acids in the C2 photorespiratory bypass. Glycerate is the product of photorespiration after being transported back into the stroma via the glycolate/glycerate transporter. Glycerate changes indicate less carbon routed through endogenous photorespiration, while also indicating similar increased levels of photorespiration in treatment B13.

	Genotype Effect		Environmental Effect	
Metabolite	B10-pT5:G1	B13-pT5:G1	pT5-B13:B10	G1-B13:B10
Sucrose	0.46	1.11	106.17	44.25
Glucose	0.99	1.02	1.33	1.29
Fructose	0.95	1.02	1.4	1.31
Ribulose/Xylulose	1.01	0.84	1.17	1.41
Xylose	0.58	1.51	2.25	0.87
Ribose	1.2	0.82	0.66	0.97
Sedoheptulose	1.01	0.84	1.17	1.41
Malate	0.59	0.77	1.63	1.25
Pyruvate	0.32	0.9	8.58	3.04
Glycerate	0.79	0.81	1.76	1.71

Fold of Change (Ratios of Group Means from Scaled Imputed Data)

	Increased (p<0.05)
	Increased (0.05<p<0.1)
	No Change (p>0.1)
	Decreased (0.05<p<0.1)
	Decreased (p<0.05)

These compounds also showed some genotypic differences, including higher ascorbate and higher levels of the reduced form of glutathione (GSH), along with lower levels of dehydroascorbate in the photorespiratory bypass line, consistent with a more favorable oxidative environment. However, this effect seemed most apparent in the higher temperature and light condition. As with the ascorbate pathway, tocopherols are also important in the maintenance of oxidative balance. Levels of the predominant species in plant leaves, alpha-tocopherol, was higher in the B13 condition. There was also an effect of genotype (**Table 4**), at least for alpha-tocopherol, which was consistent with the ascorbate and GSH observations. pT5 (C2 photorespiratory bypass) showed higher alpha-tocopherol levels in both B10 and B13 conditions relative to their G1 controls. We also saw two compounds which may be reflective of the photorespiration reactions, glycate and glycine. Glycate was lower in pT5 relative to the G1 lines, in both treatments, while glycine was higher. Free fatty acids, lyso-lipids, and monoacylglycerols (MAGs), were all more abundant in the B13 environmental condition. Membrane remodeling is an integral part of tissue growth, being necessary for cell division and expansion, explaining the increase in these detected species from condition B10 to B13. Vitamin B6 is an important component of many enzymes, as well as being reported to possess anti-oxidant activity and is linked within the salvage pathway involving the various oxidative and phosphorylated forms. Levels of these scavenging or salvaging compounds are almost always lower in the C2 photorespiration bypass (pT5) lines, especially at the cooler condition (B10), suggesting less need for salvage in those lines, and even less with lower light and temperatures. Most genotypic variances observed, comparing the photorespiration bypass line with G1 within the same treatment, were less than three-fold. Compared to the environmental difference where fold change of 10 or even 100 fold was observed.

Table 4. Genotype dependent changes in B10 and B13 for pT5 and G1

List of the genotypic related fold changes comparing the pT5 (photorespiratory bypass) to G1 (terpene sink only) background plants, in both B10 and B13 controlled environment treatments. Confidence interval of significance produced by Welch's t-test. Increased (red) and decreased (green) compounds for both environments are given, however most approaching the 3-fold threshold correlate to a more active membrane remodeling growth phenotype. Other observations worth noting are the decreases in carbohydrate metabolism in pT5 compared to G1 for both environments.

Table 4.

Super Pathway	Biochemical Name	B10 / B10_G1	B13 / B13_G1
Amino acid	sulfate*	2.14	1.42
Amino acid	tyrosine	1.93	3.01
Amino acid	asparagine	3.37	2.29
Amino acid	2-aminoadipate	1.61	2.46
Amino acid	6-oxopiperidine-2-carboxylate	1.48	2.64
Amino acid	S-methylcysteine	1.27	3.28
Amino acid	N6,N6,N6-trimethyllysine	2.08	2.07
Amino acid	N-acetylasparagine	2.02	1.2
Amino acid	arginine	1.45	2.2
Amino acid	citrulline	6.2	1.49
Amino acid	dimethylarginine (SDMA + ADMA)	2.15	1.72
Amino acid	glutamine	2.19	1.52
Amino acid	histidine	3.5	3.44
Amino acid	pyroglutamine*	2.1	1.18
Amino acid	N-delta-acetylornithine	2.8	1.84
Amino acid	N-monomethylarginine	4.81	6.4
Amino acid	spermidine	2.13	1.05
Amino acid	gamma-glutamylglutamine	2.08	1.54
Amino acid	gamma-glutamylhistidine	2.63	6.26
Amino acid	gamma-glutamylthreonine	2.23	1.62
Amino acid	gamma-glutamyltryptophan	1.87	3.92
Cofactors	alpha-tocopherol	1.37	2.06
Nucleotide	allantoin	1.76	2
Nucleotide	2'-deoxycytidine	2.79	1.02
Nucleotide	5-methyl-2'-deoxycytidine	3.2	1.56
Nucleotide	thymidine	2.95	1.2
Nucleotide	hydroxymethylpyrimidine	1.32	3.2
Hormone metabolism	indoleacetate	1.21	2.04

Super Pathway	Biochemical Name	B10 / B10_G1	B13 / B13_G1
Amino acid	shikimate	0.7	0.28
Amino acid	gamma-glutamylglycine	0.32	1
Carbohydrate	pyruvate	0.32	0.9
Carbohydrate	oxaloacetate	0.32	0.68
Carbohydrate	erythronate*	0.27	0.67
Carbohydrate	ribonate	0.23	0.71
Lipids	1-linoleoyl-GPG (18:2)*	1	0.33
Lipids	glycerophosphoglycerol	0.31	0.81
Lipids	linolenoyl-hexadecatrienoyl-glycerol (18:3/16:3)*	0.26	0.38
Cofactors	dehydroascorbate	0.08	0.4
Cofactors	threonate	0.31	0.53

0.55
0.76
1.71
1.32
1.20

Green: indicates significant difference ($p \leq 0.05$) between the groups shown, metabolite ratio of < 1.00

Light Green: narrowly missed statistical cutoff for significance 0.05<p<0.10, metabolite ratio of < 1.00

Red: indicates significant difference ($p \leq 0.05$) between the groups shown; metabolite ratio of ≥ 1.00

Light Red: narrowly missed statistical cutoff for significance 0.05<p<0.10, metabolite ratio of ≥ 1.00

Non-colored text and cell: mean values are not significantly different for that comparison

The only trend we observe is the predominance of amino acid and nucleotide metabolites, often catabolites, among the compounds which increased in the pT5/G1 comparisons. Other compounds decreased in comparison of pT5 to G1, including several carbohydrates such as pyruvate, OAA, as well as ascorbate metabolites referred to above. A sub-set of compounds were also observed and identified as having an genotype and environmental interaction as directionality of change was opposite under the two conditions (**Table 5**). This sub-set shows that environment also plays a role in the function of the bypass, and that the bypass behaves differently in altered environments.

As noted above, the environmental treatment tended to dominate the variance, while the purely genotype related differences were more subtle, and did not account for very many large fold-change differences. In the case of an n=1 comparisons, a two-fold difference is a relatively weak cut-off. The only strong trends observed were the predominance of amino acid and nucleotide metabolites, often catabolites, among the compounds which increased in the pT5/G1 contrasts. Those compounds which decreased in these contrasts included several key sugars and organic acids, as well as the ascorbate metabolites mentioned above. We also observed a sub-set of compounds which my exhibit an interaction between the environmental condition and genotype variables, in that the directionality of change was opposite under the two temperature light conditions. Again, these tended to be amino acid associated, and the magnitude of the changes were not large.

Table 5. Differential Genotype x Environment interaction metabolic reactions

List of metabolites analyzed which had contrasting response in fold change from pT5 compared to G1 in different controlled environment treatments. These compounds have a varied environmental impact on genotype. Largely observed are metabolites associated with more active growth and membrane remodeling, where in treatment B10 pT5 appears to be growing slower than G1 and quicker in treatment B13, suggesting the environment plays a larger role and impact on growth phenotype in pT5 than G1.

Super Pathway	Biochemical Name	B10 / B10_G1	B13 / B13_G1
Amino acid	tryptophan	0.89	2.17
Amino acid	3-methoxytyramine	0.67	2.02
Amino acid	saccharopine	0.82	2.05
Amino acid	pipecolate	0.87	2.02
Amino acid	2-oxoadipate	0.27	1.86
Amino acid	methionine	0.3	1.47
Amino acid	S-methylmethionine	0.47	3.15
Amino acid	N-alpha-acetylornithine	0.1	2.18
Amino acid	3-methyl-2-oxobutyrate	0.81	4.12
Amino acid	S-methylglutathione	0.44	2.61
Lipids	linolenate [alpha or gamma; (18:3n3 or 6)]	0.21	1.36
Lipids	1-palmitoyl-2-oleoyl-GPG (16:0/18:1)	1.95	0.29
Lipids	glycerol 3-phosphate	0.26	1.5
Nucleotide	allantoic acid	0.51	5.47
Nucleotide	3-methylcytidine	2.39	0.94
Hormone metabolism	indoleacetylaspartate	0.73	2.97

0.55	**Green:** indicates significant difference ($p \leq 0.05$) between the groups shown, metabolite ratio of < 1.00
0.76	**Light Green:** narrowly missed statistical cutoff for significance 0.05<p<0.10, metabolite ratio of < 1.00
1.71	**Red:** indicates significant difference ($p \leq 0.05$) between the groups shown; metabolite ratio of ≥ 1.00
1.32	**Light Red:** narrowly missed statistical cutoff for significance 0.05<p<0.10, metabolite ratio of ≥ 1.00
1.20	**Non-colored text and cell:** mean values are not significantly different for that comparison

Conclusions

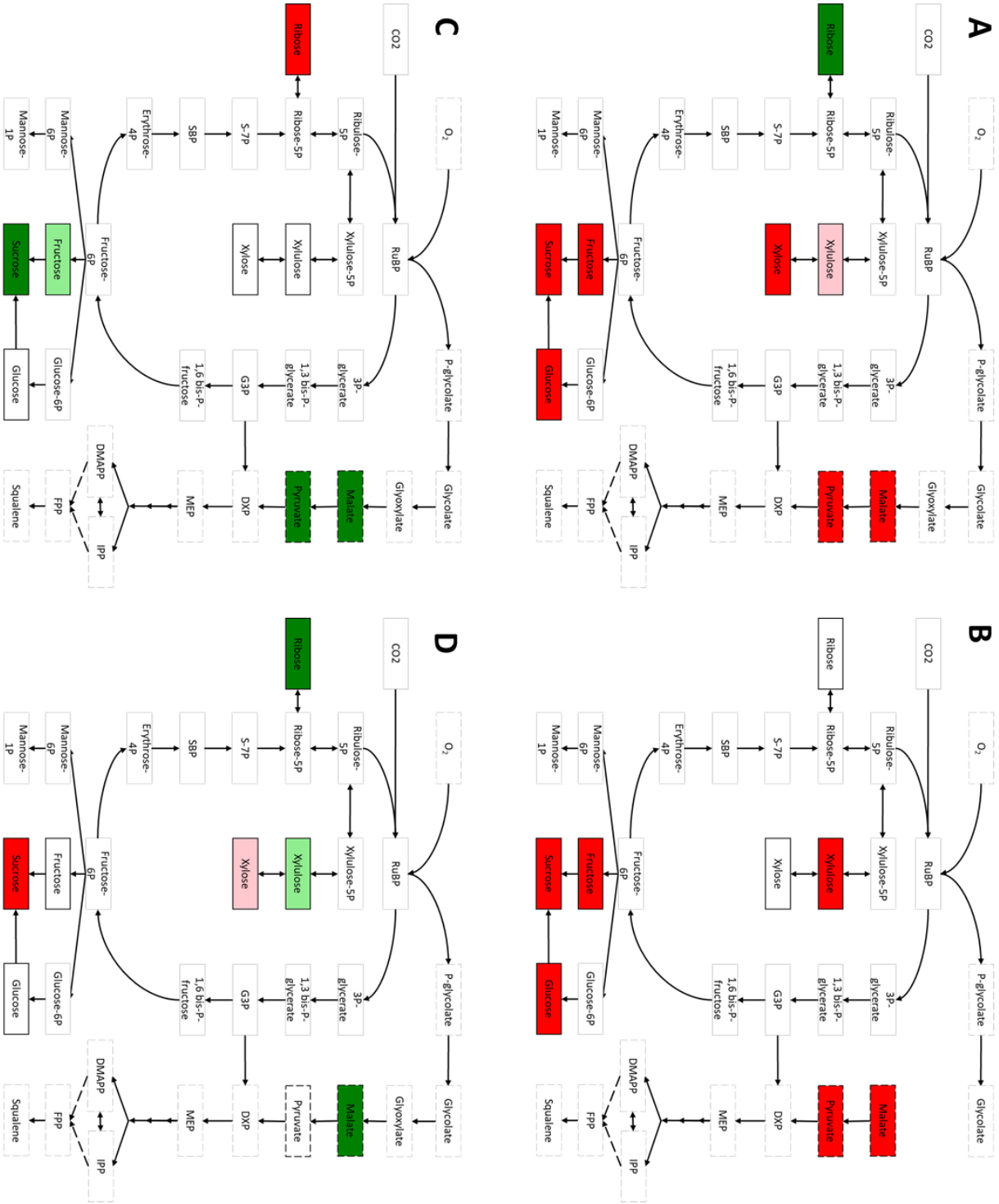
Producing a predictable phenotype relies on both a stable genotype and a controlled or at least favorable environment[59]. After high variation in yield, the C2 photorespiratory bypass tobacco lines were subjected to treatments on opposite ends of typical growth conditions for cultivated burley line tobacco to determine which variable of the phenotype equation had a greater impact on the resulting bioproduct yield. Increased photosynthesis yielded an increase in squalene bioproduct as determined by GC/MS analysis, but the relative gain observed in the C2 photorespiratory bypass plants diminished under increasing photosynthesis and photorespiratory conditions. This pathway is limited in flux capacity, limiting its efficacy as photorespiration rates increase, and contributing less to the overall carbon allocated to terpene biosynthesis as photosynthesis rate increases. Overall, other than metabolomics associated with active growth and remodeling accounting for the most variance between comparisons, a small genetic variance was observed between genotypes. It suggests that while the fixation rate remains not significantly different, the C2 photorespiratory bypass may experience a more conducive fixation environment and require less active balancing of reduction potential and oxidation energies. Furthermore, individual elements of the design such as malate synthase may contribute to sink strength under low fixation conditions. Sugar biosynthesis is decreased as carbon is allocated to the strong terpene sink in both genotypes (**Figure 8**). Pyruvate is higher in condition B13 as expected from both photosynthesis and the bypass design, but only in condition B10 has a strong genotype effect, suggesting increased consumption of pyruvate by the bypass pathway to construct DXP. The overall scheme shows a partition in primary metabolism to feed the terpene sink, with the photorespiratory bypass providing some additional gain but mitigated when photosynthesis rate is increased. Due to this metabolic shift in carbon flux to the terpene biosynthesis sink, other

strategies without a direct coupling to carbon fixation through energy, reductant, and carbon source may provide a more stable way to direct primary metabolism carbon to terpene biosynthesis without the strong environmental impacts as the C2 photorespiratory bypass pathway.

Figure 8. Carbon Flux Models for Genotype and Environmental Impacts on pT5 and G1 in B10 and B13

Simplified model of key compounds of the Calvin Benson cycle, C2 photorespiratory bypass pathway, MEP pathway, and engineered terpene sink to produce squalene in the chloroplast stroma. Black outlined metabolites were analyzed in metabolomics, gray are model place holders. Solid outlined metabolites are of C3 fixation and sugar metabolism. Dashed outlined metabolites are of C2 redirection and MEP pathway to squalene. Values and significance for flux models can be read from Table 3. Model showing metabolic flux changes for (**A**) Environmental impact on the carbon metabolism of pT5, comparing B13 to B10; (**B**) Environmental impact on the carbon metabolism of G1, comparing B13 to B10; (**C**) Genotypic associated variation of pT5 compared to G1, in controlled environment B10; and (**D**) Genotypic associated variation of pT5 compared to G1, in controlled environment B13.

Figure 8.


A
B
C
D
CO2
O2
RuBP
P-glycolate
Glycolate
Glyoxylate
Malate
Pyruvate
Ribulose-5P
Xylulose-5P
3P-glycerate
1,3 bis-P-glycerate
Ribose
Ribose-5P
Xylulose
S-7P
Xylose
SBP
G3P
1,6 bis-P-fructose
Erythrose-4P
DXP
MEP
Mannose-6P
Mannose-1P
Fructose-6P
Fructose
Sucrose
Glucose-6P
Glucose
DMAPP
IPP
FPP
Squalene

IV. CONSTRUCTION, VALIDATION, AND EVALUATION OF DIRECT C5 CONVERSION TO DXP

Introduction

The environmental impact on terpene production is known to be important. Even in engineered constitutively driven systems this is true. In photosynthetic platforms the synthesis of both sugars for primary metabolism and terpene biosynthesis rely on G3P produced from the Calvin Benson cycle[13]. Both terpenes and central metabolism require carbon, thus promoting the idea to engineer a photorespiratory bypass metabolic pathway to capture some carbon lost during photorespiration and direct it into terpene bioproduct. Some gains were made with this strategy, however the environmental impact variable on photorespiration rates compared to photosynthesis made yield prediction of these plants difficult. We seek to engineer a more direct route utilizing pentose sugars and pentose phosphates converted directly to DXP, which greatly reduces the kinetic steps to get to the MEP pathway while producing a C5 compound without evolving previously fixed carbon in the process. Gains have been observed from this design, along with coupling these designs with other modifying genetic cassettes. This route may also offer a possible design route for a temporal or possibly induced promoter driven system, allowing for rapid conversion of sugars and biomass to bioproduct pre-harvest.

Due to the intrinsic ties between photosynthesis and the MEP pathway, terpene biosynthesis is already highly reliant on light reactions for energy and reductant and the Calvin Benson cycle for carbon[2]. A partial photorespiration bypass works to capture some carbon lost to photorespiration, but is reduced in efficacy when photosynthesis rates increase. For these reasons this design is highly impacted by the environment, and sees diminishing gain when growth

conditions are optimized. Furthermore, the current C2 photorespiratory bypass pathway re-routes p-glycolate, a two carbon compound, while evolving $2CO_2$ in the process. The goal is to route carbon to DXP, the first committed step in the MEP pathway, which is a five carbon compound. Many other five carbon compounds exist within the plant such as isoprene and pentose sugars[64, 65]. The later, along with their phosphorylated forms, make up the back bone of primary metabolism, the pentose phosphate pathway. These C5 compounds are isomerized and phosphorylated and rearranged to supply various primary metabolic functions with a carbon source. For this reason they are ample and typically relatively easy to convert into other pentose sugars or pentose phosphates[64]. Siphoning carbon from the pentose phosphate pool would be expected to cause less unexpected consequences within primary metabolism as other C5 compounds will be rearranged to compensate the loss.

DXP (Deoxy-Xylulose-Phosphate) is also a C5 compound. The reactions to get from pentose sugars and sugar phosphates directly to DXP however are only recently being discovered through mutant study and microbial biotechnology[66, 67]. Plants typically use the pentose pools and carbon sources for other reactions, where the pentose sugars and phosphates largely are only acted upon by kinases/phosphatases, transporters, and isomerization enzymes[64]. To get to DXP requires oxidation. In one scheme, the gene *YajO* from *E. coli* is a recently discovered 1-deoxyxylulose-5-phosphate synthase. This catalyzes the conversion of ribulose 5-phosphate (Ru5P) or xylulose 5-phosphate (Xu5P) to 1-deoxy-D-xylulose 5-phosphate (DXP)[68]. Putative study believed *YajO* to act on xylulose to produce 1-deoxyxylulose which was then phosphorylated by *XylB*, a xylulose kinase[67]. This provides a direct route from pentoses to terpenes in plants, with the addition of only the *YajO* cassette and function of the endogenous xylulose kinase enzyme. Another scheme, utilizing the *ribB* mutant also discovered in *E. coli* rerouting Ribulose-5P or

Xylulose-5P directly to DXP, in one step. The efficiency of this system relies on its simplicity, and thus could be controlled simply by enzyme titer.

The overall scheme of the C5 to C5 direct conversion from a C3 plant to downstream terpene sink is laid out in **Figure 9**. Permutations of this design for us in a binary plasmid along with other cassettes identified as additive to the bioproduct yield were also designed and constructed (**Figure 10**). The C5 to C5 engineered route seeks to increase carbon flux directly into the MEP, but does little to increase efficacy of the MEP itself, and instead seeks to produce DXPS without reliance on G3P and pyruvate. The pull of ctpFPPS and ctpSQS create the strong constitutive downstream terpene sink required to produce squalene at extractable titers[2]. When combined with other additive cassettes such as Zhao Cheng's truncated-oleosin "oil droplet" for increase storage of the squalene synthesized, and SBPase to try to increase fixation strength, there is a slight gain (FSTR, FSSR). Other permutations including DXPS may be introduced to increase the carbon allocation to MEP, as increasing DXPS titer leads to a more competitive recruitment of G3P and pyruvate to terpene biosynthesis.

Materials and Methods

Plasmid Construction

Coding sequences for RibB and YajO along with a high efficintcey xylulose kinase, XylB, were synthesized by IDT from *N. tabacum* codon optimized *E. coli* gene sequences for the construction of these binary vectors.

Figure 9. Design of the C5 to DXP direct carbon flux through RibB or YajO

Design of C5 to C5 engineered carbon flux strategies (red) utilizing RibB, and YajO (with or without XylB) to convert pentose phosphates to DXP, omitting the condensation reaction of G3P and Pyruvate yielding a loss of carbon in the form of evolved CO_2.

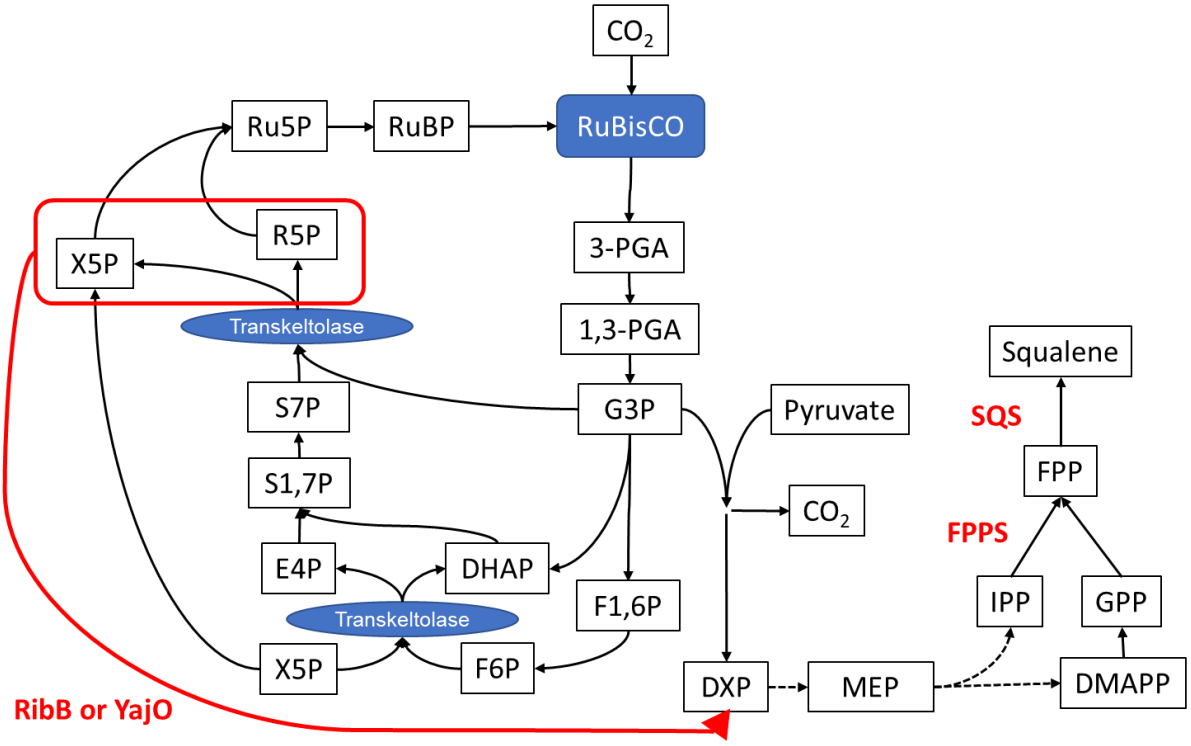
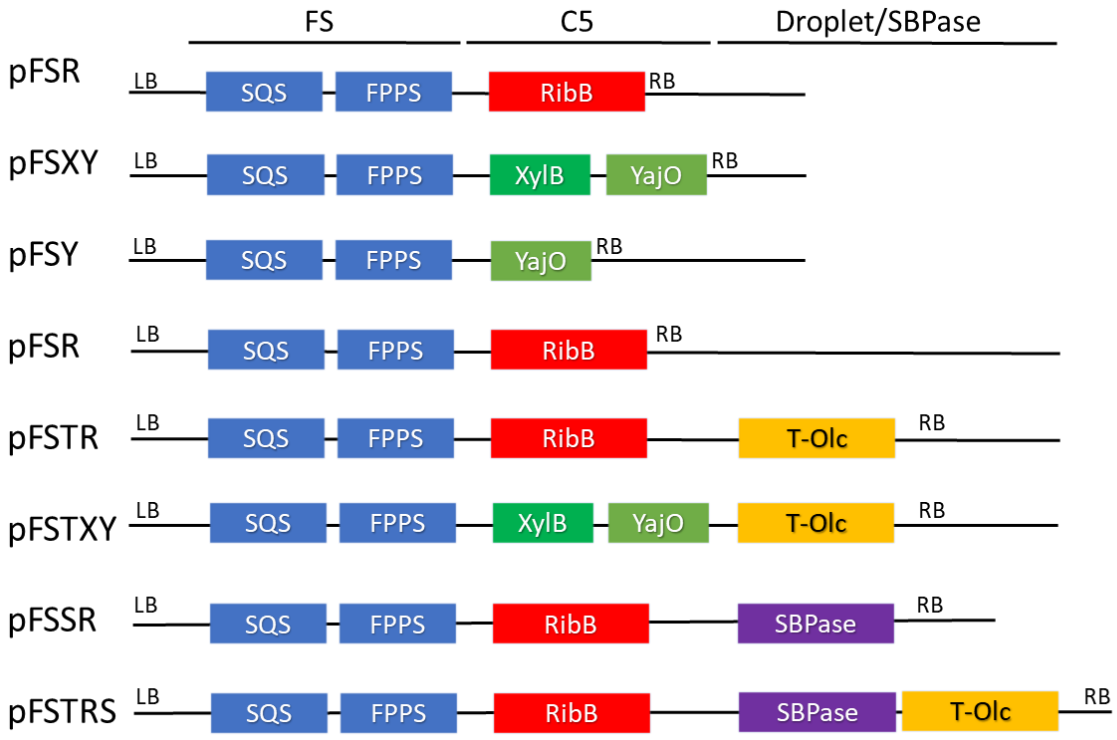


Figure 10. Design of C5 to DXP with terpene sink binary plasmids

Design of terpene production binary plasmid vector design, utilizing the chloroplast transit peptide in each cassette, each contains a FPPS+SQS terpene sink, a C5 to C5 routing, with and without coupling to T-oleosin droplet or SBPase.



The C5 to C5 constructs, were assembled using the Gibson self-assembly method and cloning in DH5α and 10-β competent cell lines. Then transformed into *A. tumefaciens* GV1301 for plant transformation.

The YajO and YajO with optimized XylB constructs were also assembled using the Gibson self-assembly method, along with traditional restriction enzyme digestion to linearize and truncate a back bone from the pTerpene series of plasmids. Full cloned cassettes from amplicons with adjoining overhangs were assembled along with the FPPS and SQS cassettes fragment cloned from G1 and the linearized backbone to produce complete binary plasmids, recovered by transformation and selection of DH5α *E. coli* competent cells. The genetic cassettes for ctpFPPS and ctpSQS were cloned directly from G1 in one fragment from overlapping amplicons. After the CaMV driven NeoR poly-A terminated antibiotic resistance cassette which runs in reverse order from the rest of the T-DNA, the first genetic cassette from the left border 5'-3' is the truncated squalene synthase from yeast fused to a 5' chloroplast transit peptide from the small subunit of *A. thaliana* RuBisCO. It is driven by the Pcv promoter from the Cassava vein mosaic virus, and is terminated by Tnos as all the cassettes other than selection cassettes are in this construct. A chloroplast transit peptide fused farnesyl pyrophosphate synthase is the next cassette driven by the p35S promoter. In the construct with the xylulose kinase, *E. coli* XylB is driven by Pcv, and preceeds the YajO cassette. This last cassette features a chloroplast transit peptide fused to the YajO coding sequence, and is also terminated by Tnos.

RibB constructs were built using similar methods, along with permutations with cassettes for SBPase and T-oleosin droplet to further increase squalene yield. The RibB gene is driven by the Pcv promoter and terminated by Tnos. SBPase and T-oleosin are driven by p35S and Gateway

35S respectively, and also terminated by Tnos. Permutations of the C5 to C5 binary plasmid constructs T-DNA are shown in (**Figure 11**).

Tobacco Transformation

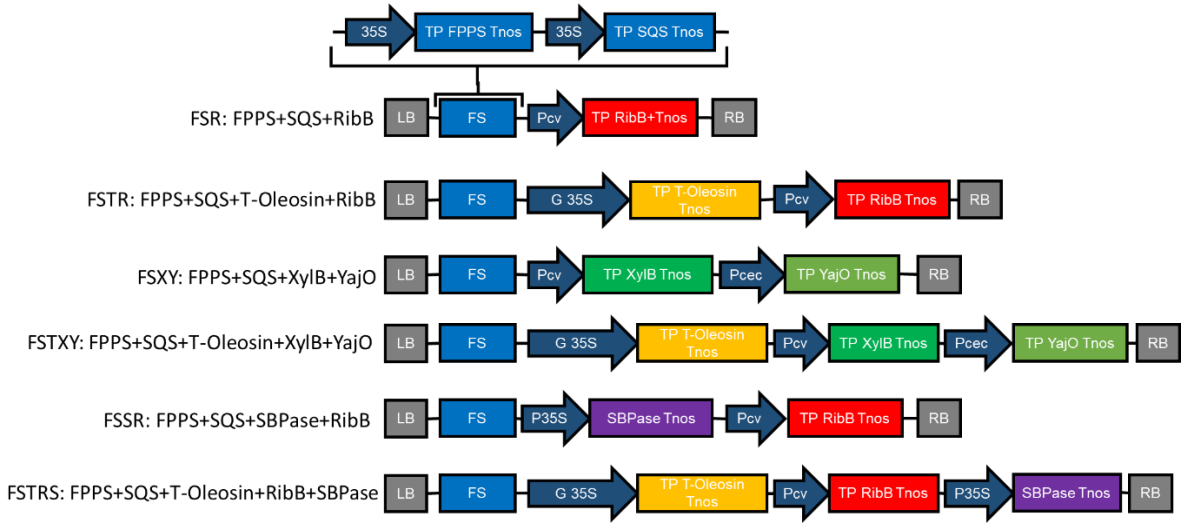
Line 1068 WT burley tobacco seeds were germinated on sterile MS media in a parafilm sealed petri dish under a 16hr light cycle at 24°C for approximately 10 days, when the cotyledons were fully emergent and the radicle had begun to make first hairy roots. From this stage tobacco seedlings were placed into a sterile 500ml glass jars, sealed, and placed back into the light room to grow.

Confirmed binary plasmids were isolated from the *E. coli* cloning LB selection plates. These were used to transform line GV1301 Agrobacterium tumefaciens, via heat shock transformation. After confirmation by restriction enzyme digestion and PCR of plasmid isolations derived from colonies growing on YEP selection media, an individual colony was picked to grow a liquid culture for to perform plasmid isolation for final sequencing and then again for transformation of tobacco leaf material. Culture was grown to 0.8 OD_{600} in YEP liquid media with selection antibiotics, then bacteria were pelleted using centrifugation, and re-suspended in a ½ MS-0% sucrose liquid media at an OD_{600} of 0.6 pre-transformation.

Sterile tobacco leaf tissue of juvenile plants (2-3 weeks old) was excised, cut to approximately $1cm^2$, and added to the agrobacterium suspension then allowed to shake for 1 hour at 200rpm in the orbital shaker. The agro-suspension culture was then decanted off of the tobacco explants, and discarded. The explants were then briefly washed using sterile H_2O to reduce the agrobacterium overgrowth during co-cultivation.

Figure 11. Constructed T-DNA of binary plasmids showing adjoined fragments

Constructed permutations C5 to C5 to Terpene design into binary plasmids, showing individual fragments used in the self-assembly. Use of the same amplicons to build multiple permutations ensures elements are comparable and streamlines molecular work prior to transformation of tobacco. All promoters are constitutive viral promoters and Tnos is used for transcriptional termination in all cassettes.



Explants were then removed from the suspension vessel, laid out inside the laminar flow hood on sterile paper towels, and allowed to dry for 20-30min before being placed in to petri dishes. This helped with agro overgrowth and improved recovery rate in transformants. Next, transformed explants were placed abaxial side down on MS media for co-cultivation in the dark. After 2 days, explants were washed three times with sterile H_2O and transferred to MS media with antibiotic selection for two weeks under the 16hr light cycle, supplied by multiple T5 bulbs. Following initial transgene selection and bacterial selection, explants were transferred to new selective media every 2 weeks, until new leaf primordia formed indicating a new apical meristem to excise from the explant. In some transformation batches, additional supplementation of the media was required to obtain physiological development such as $1mg*L^{-1}$ BAP (6-Benzylaminopurine) with and without $0.1mg*L^{-1}$ NAA (Naphthalene acetic acid). Some transformation batches were given 1 or $2mg*L^{-1}$ of IAA (Indole acetic acid) to aid in rooting, but generally this additional step was done to increase throughput and was not necessary for root formation. After sub-culturing and regenerating shoots and roots, tobacco explants were removed from *in vitro* culturing and placed in domed seedling starter trays under T5 bulbs at 16hr light to acclimate to soil. After 2 weeks of soil growth, the humidity dome was removed to allow hardening off. From this point, plants were transferred to 1 quart plastic pots and allowed to grow under greenhouse HPS (high pressure sodium) supplementary lights and full sun. Last, these T_0 plants were transferred to 5 gallon pots, and are able to be collected from for squalene quantification after the 7^{th} fully formed leaf is developed and matured.

Squalene Quantification

Leaf tissue samples were monthly excised for squalene extraction. As within prior study, the material harvested for squalene yield analysis came from the first mature leaf from the top. The

0.5g fresh leaf samples were collected from mature unshaded leaves and macerated with a glass pestle after using liquid N_2 to flash freeze samples in 50ml falcon tubes, producing a powdered leaf material. Samples were then extracted using hexane by shaking the powdered leaf material and 3ml hexane in an orbital shaker for 2.5hrs. Cedrene was added at a final concentration of 15 ppm to serve as an internal control for the extraction. Cedrene being a terpene gives it similar extraction and degradation properties to squalene but being a sesquiterpene is half the mass and therefore can be discerned by GC/MS analysis. The extracts were then purified by passing through a 3.4cm silica (size 60 = 40-63μm) column in a glass pasteur pipette plugged with glass wool at the bottom and a thin layer of sodium sulfate ($NaSO_4$) on top. Half (1.5ml) of the hexane extract was pipetted through the silica chromatography column and collected in a glass vial. Then the column was washed with hexane until the final collection volume reached 6ml, allowing for all squalene to have passed through into the glass vial. (A visual marker of this completion is the yellow/orange band of beta-carotene, a tetraterpenoid, reaching the glass wool.) From here the 6ml final collection volume was given a nitrogen headspace and placed in -20°C. From these vials, 1ml was put into GC/MS amber vials and was run on the GC/MS. GC/MS analysis was performed using the GCMS-QP2010SE (Shimadzu). One microliter of sample was injected into the GC/MS using an AOC-20i auto-sampler in 30:1 split mode (injector 220°C) onto a ZB-5MSi fused silica capillary column (30 m × 0.25 mm × 0.25 μm thickness). The initial oven temperature was initially set to 40°C with a 30sec hold, then ramped up to 140°C at a rate of 30°C/min, then to 150°C at a rate of 5°C/min, then to 280°C at a rate of 25°C/min, and finally up to 300°C at a rate of 5°C/min with a 2min hold. Helium was used as the carrier gas. The ion source was set to 230°C and the interface was 280°C. Squalene quantification was performed using selected ions. Peak identification of the compound was performed using direct comparison of the sample mass

chromatogram with those of commercially available standard compounds. The quantitative calculations of squalene concentration were based on the peak area ratios relative to those of the standard. To ensure the precision of the measurements, one sample from each extraction batch was run as a technical replicate (from same 6ml vial and ran on GC/MS during each processing batch). The GC/MS squalene yield in ppm was then converted to µg/g fresh weight.

Metabolome Methods

Three samples from T_1 FSR (ctpFPPS+ctpSQS+ctpRibB) and T_1 FS (ctpFPPS+ctpSQS) each were sent to Metabolon for metabolomics analysis. These plants were mature and pre-flowering when leaf material was collected, lyophilized, and sent out for analysis. Increased and decreased metabolites were discovered using techniques described in Chapter 2. Significance of metabolic variation was determined by Welch's t-test, and given in fold change in (**Table 6**).

Photosynthesis Measurements

Photosynthetic activity was measured for engineered genotypes in the greenhouse, using a LI-COR® LI-6400 portable photosynthesis system from LI-COR® Biosciences. The LI-6400 yields measurements of photosynthesis and transpiration based on the differences in CO_2 and H_2O in an air stream that is flowing through the leaf cuvette. The sealed chamber is able to make a measurement between the inlet modified atmosphere and the return line leading out of the chamber. This allows us to measure the amount of assimilated carbon and determine active photosynthetic rate. A total of 5 biological replicates for each genotype were used with 3 technical replications for each biological replicate.

Table 6. Metabolite changes of FSR compared to FS

Tables of increased metabolites (left, red) and their given fold changes for metabolites greater than 2 fold change, and significantly decreased metabolites (right, green) for the comparison FSR to FS. Light and dark colors of increase and decrease indicate statistical cutoff, between $p \leq 0.05$ and $0.05 < p < 0.10$. Largely the differences yielded increase in some organic acids in carbohydrate metabolism associated with the Calvin Benson cycle and pentose phosphate pathways and alteration in lipid content and amino acids attributed to growth variation caused by RibB activity.

Table 6.

Super Pathway	Biochemical Name	FSR / FS
Amino acid	kynurenate	2.43
Amino acid	alpha-ketoglutaramate**	2.11
Amino acid	gamma-glutamylglutamate	3.18
Carbohydrate	glycerate	2.32
Carbohydrate	erythronate*	2.37
Carbohydrate	ribonate	2.49
Carbohydrate	ribulonate/xylulonate*	2.02
Lipids	linolenate [alpha or gamma; (18:3n3 or 6)]	2.43
Lipids	3-hydroxybutyrate (BHBA)	4.09
Lipids	1-palmitoyl-2-oleoyl-GPA (16:0/18:1)	2.19
Lipids	1-palmitoyl-2-linolenoyl-GPA (16:0/18:3)*	2.02
Lipids	1-palmitoyl-GPA (16:0)	2.7
Lipids	1-stearoyl-GPA (18:0)	2.17
Lipids	1-linoleoylglycerol (18:2)	2.09
Lipids	1-linolenoylglycerol (18:3)	4.13
Lipids	2-linoleoylglycerol (18:2)	2.87
Cofactors	dehydroascorbate	5.9
Nucleotide	adenosine	2.69
Nucleotide	xanthosine	2.15
Secondary metabolism	normetanephrine	2.03
Secondary metabolism	chlorogenate	2.86
Secondary metabolism	cryptochlorogenic acid	2.81
Secondary metabolism	coumaroylquinate (2)	2.19
Secondary metabolism	coumaroylquinate (4)	6.58
Secondary metabolism	3-hydroxy-3-methylglutarate	2.78

Super Pathway	Biochemical Name	FSR / FS
Amino acid	S-carboxymethyl-L-cysteine	0.15
Amino acid	tryptophan	0.16
Amino acid	3-(4-hydroxyphenyl)propionate	0.45
Amino acid	O-sulfo-L-tyrosine	0.3
Amino acid	N-acetyltryptophan	0.39
Amino acid	asparagine	0.35
Amino acid	pipecolate	0.37
Amino acid	S-methylmethionine	0.22
Amino acid	citrulline	0.45
Amino acid	histidine	0.18
Amino acid	N-delta-acetylornithine	0.33
Amino acid	2,3-dihydroxyisovalerate	0.43
Amino acid	beta-hydroxyisovalerate	0.48
Amino acid	putrescine	0.23
Carbohydrate	aconitate [cis or trans]	0.44
Lipids	12,13-DiHOME	0.47
Lipids	1-oleoyl-2-linoleoyl-GPC (18:1/18:2)*	0.4
Lipids	stearoyl-linoleoyl-glycerol (18:0/18:2) [1]*	0.35
Lipids	oleoyl-linoleoyl-glycerol (18:1/18:2) [2]	0.45
Nucleotide	allantoin	0.05
Nucleotide	urate	0.3
Nucleotide	xanthine	0.35
Nucleotide	pseudouridine	0.39
Nucleotide	thymidine	0.39
Nucleotide	uridine-2',3'-cyclic monophosphate	0.47
Nucleotide	beta-alanine	0.48
Nucleotide	uridine 3'-monophosphate (3'-UMP)	0.19
Hormone metabolism	indoleacetate	0.29
Hormone metabolism	indoleacetylaspartate	0.26
Hormone metabolism	2-oxindole-3-acetate	0.41
Hormone metabolism	indole-3-carboxylic acid	0.46
Secondary metabolism	2,4,6-trihydroxybenzoate	0.42

0.55	**Green:** indicates significant difference ($p \leq 0.05$) between the groups shown, metabolite ratio of < 1.00
0.76	**Light Green:** narrowly missed statistical cutoff for significance 0.05<p<0.10, metabolite ratio of < 1.00
1.71	**Red:** indicates significant difference ($p \leq 0.05$) between the groups shown; metabolite ratio of ≥ 1.00
1.32	**Light Red:** narrowly missed statistical cutoff for significance 0.05<p<0.10, metabolite ratio of ≥ 1.00
1.20	**Non-colored text and cell:** mean values are not significantly different for that comparison

Results and Discussion

The C5 pathway increases squalene

As aforementioned, considering the significant limitation of C2 redirection, we decided to explore the possibility of directly channeling C5 compounds from the Calvin Benson cycle to DXP to produce terpene molecules. The strategy could have at least two advantages. First, this strategy will by-pass the condensation reaction of G3P and pyruvate, which loses one carbon from six carbon to form the C5 DXP. The direct conversion of C5 carbon into DXP thus represents a higher carbon efficiency alternative to produce terpene precursors as compared to the MEP pathway. Second, the direct channeling of C5 carbon into terpene precursors could overcome the tight regulation at DXP biosynthesis and directly channel larger flux from the C5 pools of the Calvin Benson cycle and pentose phosphate pathways into terpene biosynthesis. The design thus represents a new route for photosynthetic output to directly synergize the 'source' and 'sink' between photosynthesis and an 'unconventional' product pool. This strategy also allows us to evaluate how to repartition carbon between primary and secondary metabolisms from a fundamental perspective.

In order to implement the new C5 design, the synthetic pathways as shown in Figure 9 will be implemented, where the Xylulose-5-Phosphate (Xu5P) will be directly converted into DXP by a mutant enzyme RibB from *E. coli*. Considering the rapid conversion between Ribulose-5-Phosphate (Ru5P) and Xu5P, the pathway basically can convert the pentose sugars and pentose phosphates from the Calvin Benson cycle into DXP for terpene biosynthesis. In addition to G3P, this represents a new outlet for photosynthesis, which might be less regulated and competed for than other major primary metabolites. The pathway implementation was carried out in tobacco

with the plasmid design as shown in **Figure 10**. The RibB enzyme coding sequence was fused in frame with *A. thaliana* RuBisCO small subunit chloroplast transit peptide coding sequence and transformed together with ctpFPPS and ctpSQS cassettes to enhance terpene biosynthesis. All enzymes were compartmented to the chloroplast stroma to create 'source' and 'sink' synergy with a strong squalene carbon sink. Previous studies have demonstrated that squalene can be accumulated to up to 2.8mg/g FW, which makes it a perfect alternative carbon sink. RibB will directly link photosynthetic C5 carbon with terpene biosynthesis for squalene carbon sink.

The plasmid were transformed into tobacco with agrobacterium transformation. The transformants were confirmed with Genomic DNA PCR for the integration of transgenes in genome. As this can also yield false positives from the transformation and co-cultivation, the gene expression was confirmed using the expression of RibB mRNA *in planta*. The reverse transcription (RT)-PCR was carried out from total RNA of transgenic tobacco and confirmed the expression of transgenes RibB and FPPS, along with RuBisCO expression for reference as shown in **Figure 12**. The positive transgenic plants were processed for squalene analysis and the T_0 plants were evaluated at different developmental stages as seen in **Figure 13**.

The results highlighted the effectiveness of C5 redirection. As shown in Figure 13, squalene yield is higher at different developmental stages for selective FSR lines with FPPS, SQS, and RibB as compared to that of FS lines with FPPS and SQS only. As shown in **Figure 14A**, the results from T_1 generation confirmed the T_0 data, where the FSR lines can reach nearly 1.8mg/g FW of squalene, but the highest yield for FS lines are around 1mg/g FW. The results again confirmed that RibB can effectively channel Xu5P/Ru5P directly to DXP to increase the squalene yield. Overall, the results highlighted the effectiveness of C5 redirection in increasing carbon partition into terpene from photosynthesis.

Figure 12. RT-PCR of RibB and FPPS expression for T_0 event confirmation

RT-PCR of RibB and FPPS expression for T_0 line events transformed with pFSR. Relative quantification to RbcS (RuBisCO subunit) expression. Line FSR 1&2 are positive for RibB expression and FPPS expression.

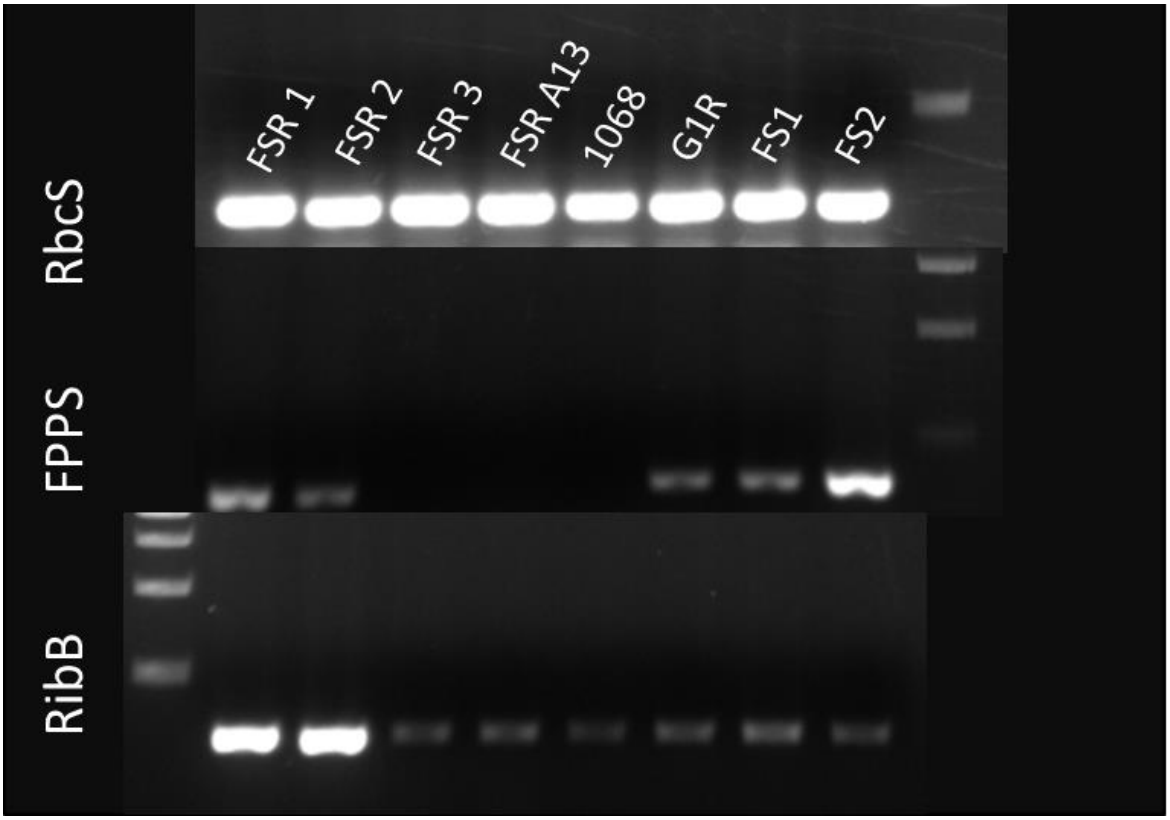


Figure 13. Squalene Quantification of T_0 FSR plants

Squalene quantification results from GC/MS analysis, extracted from monthly leaf tissue samples of ctpFPPS+ctpSQS+RibB (FSR) and ctpFPPS+ctpSQS (FS) only. RibB transformed lines were able to increase squalene yield a full fold in one event at maturity. RibB plants are able to produce higher squalene than FS alone, once they reach maturity and before flowering. After flowering all genotypes experience a drop off and sink strength changes to the developing inflorescence. A growth and productivity lag may also be occurring, as fixation pentoses are being utilized for terpene biosynthesis.

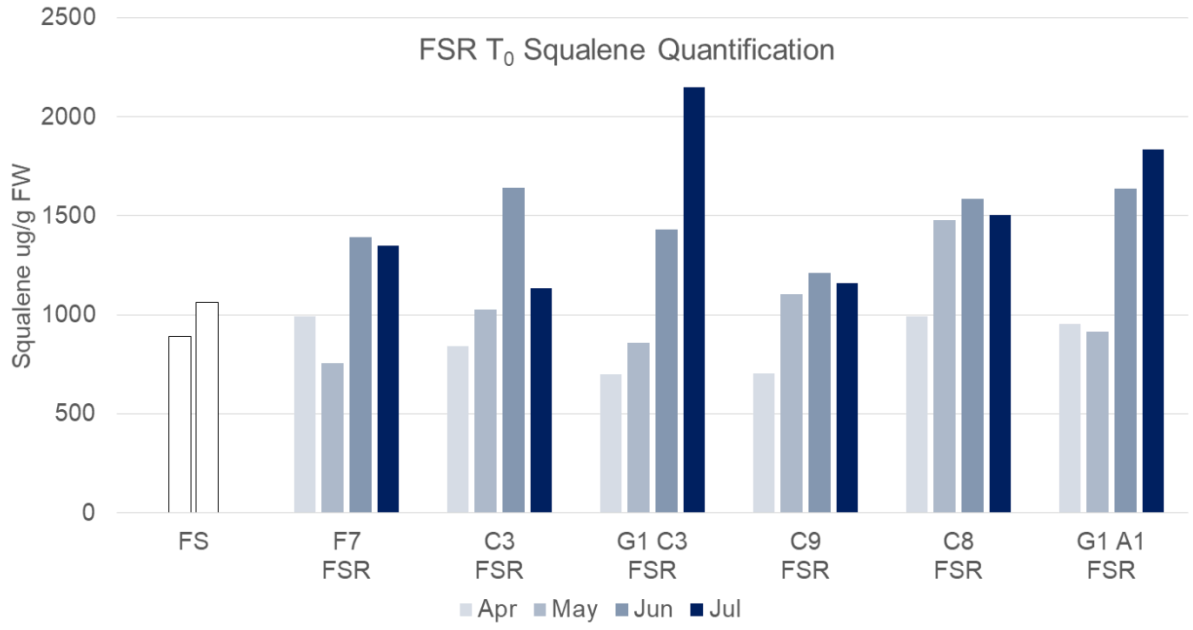
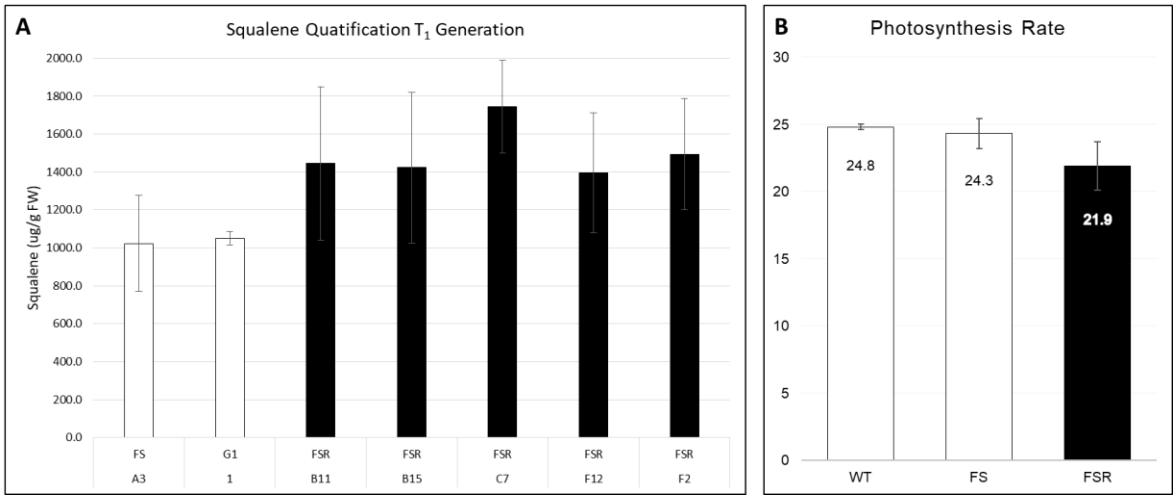


Figure 14. Squalene Quantification and Photosynthesis Rates for T_1 generation plants

(**A**) Squalene yield for T_1 plants in ug/g Fresh Weight, showing ctpFPPS+ctpSQS in all line make squalene, that FS (also T_1) and homozygous line G1 are comparable, and that plant lines with RibB mutant enzyme to route Ru5P to DXP directly see a 40-60% increase in squalene yield. (**B**) Shows the relative photosynthetic rate measured by the LiCor LI-6400 at peak mid-morning photosynthesis, along with a small but significant reduction in FSR lines, suggesting pentose and pentose phosphates utilized by the Calvin Benson cycle are being rechanneled into terpenes via RibB enzyme.



The C5 redirection slightly decreased carbon fixation

One of the key fundamental question lies in if the C5 redirection impacts photosynthesis rate or not. Our previous studies have revealed that C2 redirection did not change photosynthesis rate significantly. However, the C5 redirection involves direct rechanneling of pentose phosphate intermediates of the Calvin Benson cycle, which could lead to lower RuBP (Ribulose Bisphosphate) available for RuBisCO, resulting in reduced fixation. The results could be a reduced carbon fixation rate. The photosynthesis carbon fixation as shown in **Figure 14B** has demonstrated that the FSR lines actually have lower photosynthesis rate as compared to FS lines. The results highlighted that alternative strategies needed to channel more carbon to the Xu5P and Ru5P pools to provide sufficient C5 carbon for photosynthetic carbon fixation. Without replenishing the C5 carbon required for fixation from other pentose phosphate pathway channels the fixation rate could be compromised under limited fixation rates.

The C5 redirection changed the carbon flux to terpene

In order to further understand the mechanisms of C5 redirection and to guide the further design of efficient 'source' and 'sink' synergy for photosynthetic terpene production. We carried out metabolomics analysis to compare the FSR and FS lines. The metabolomics analysis highlighted three discoveries relevant to the C5 redirection as shown in **Figure 6**.

First, the effective C5 redirection has led to the increase of pyruvate level. C5 redirection by-passes DXPS to produce DXP, which leads to a higher IPP/DMAPP flux and squalene yield. This alternative route could end up reducing the flux for condensation of pyruvate and G3P toward DXP. In fact, a higher IPP/DMAPP level could end up inhibiting the DXPS activity, leading to lower activity for condensation of G3P and pyruvate. The lower DXPS activity could lead to the

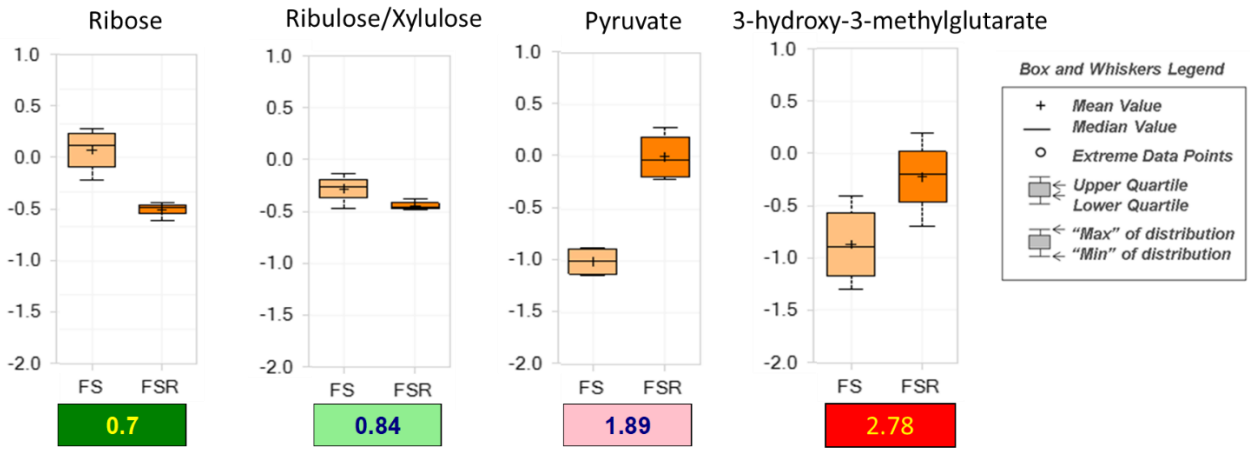
accumulation of pyruvate, as shown in the metabolomics analysis. The accumulation of pyruvate indicates the effectiveness of C5 redirection. Moreover, additional strategies could be used in future studies to restore DXPS activity to lead to synergy on DXP flux increases.

Second, metabolomics analysis revealed a general reduction of C5 intermediates, which could be an indirect effect of the larger pentose phosphate pathway rearrangement compensating for reduced C5 sugar phosphate needed for fixation. As shown in **Figure 15**, the ribose level *in planta* has decreased by 30%, while ribulose and xylulose showed a slight decrease of 16%. The decrease in C5 sugar actually correlated with the lower photosynthetic carbon fixation rate, which indicated a lower C5 sugar phosphate intermediate pool in the Calvin Benson cycle. The results indicated that additional strategies are needed to mitigate the impact of C5 redirection on C5 sugar phosphate. Two mitigation strategies could be carried out. First, we could design pathways to channel more carbon from pentose phosphate pathway and other intermediates to C5 sugar pool and further enhance the phosphorylation to produce more intermediates for the Calvin Benson cycle. Second, we could accelerate the regeneration of sugar phosphate with overexpression of SBPase in the Calvin Benson cycle. Overall, considering the lower C5 sugar pool and the lower photosynthetic carbon fixation, we should identify alternative strategies to enhance the C5 intermediate in the Calvin Benson cycle to further improve terpene yield without compromising photosynthesis.

Third, we have found an increase in 3-hydroxy-3-methylglutarate (HMG) in FSR lines. The increase could be associated with the higher IPP and DMAPP intermediate *in planta*, which could lead to the lower efficiency of both MEP and MVA pathways. The lower MVA flux also could lead to a higher 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) level, which could lead to the accumulation of 3-hydroxy-3-methylglutarate.

Figure 15. Box and whiskers plots of key metabolites in C5 and MEP for FSR:FS

Box plot and data from heat map of key metabolites profiled in this study, ribose, ribulose/xylulose which are targeted pentose pools, pyruvate previously required for to synthesize DXP, and the terpenoid 3-hydroxyl-3-methylglutarate (HMG) used in the MEP to produce DMAPP and IPP. Comparisons for heat map are given FSR compared to FS. Red and green shaded cells indicate $p \leq 0.05$ (red indicates that the mean values are significantly higher for that comparison; green values significantly lower). Light red and light green shaded cells indicate $0.05 < p < 0.10$ (light red indicates that the mean values trend higher for that comparison; light green values trend lower). A small but statistically significant decrease is observed in the amount of ribose and ribulose/xylulose in RibB plants. Also an small increase in pyruvate in observed, and a larger more significant increase in HMG content, showing increased carbon flux through the MEP toward DMAPP and IPP.



0.55	**Green:** indicates significant difference ($p \leq 0.05$) between the groups shown, metabolite ratio of < 1.00
0.76	**Light Green:** narrowly missed statistical cutoff for significance $0.05 < p < 0.10$, metabolite ratio of < 1.00
1.71	**Red:** indicates significant difference ($p \leq 0.05$) between the groups shown; metabolite ratio of ≥ 1.00
1.32	**Light Red:** narrowly missed statistical cutoff for significance $0.05 < p < 0.10$, metabolite ratio of ≥ 1.00
1.20	**Non-colored text and cell:** mean values are not significantly different for that comparison

The future of C5 redirection for terpene engineering

Overall, the results from the study suggested two aspects. On one side, the C5 redirection strategy is effective and could lead to a higher accumulation of terpene products, providing the downstream terpene biosynthesis is efficient. On the other side, the C5 redirection could lead to lower Calvin Benson cycle intermediates and photosynthetic carbon fixation. The new terpene flux could also block the existing terpene biosynthesis pathway. Both lower photosynthesis rate and the blocking of existing terpene biosynthesis capacity will make the strategy less efficient in increasing terpene yield. Even though the combined effects still lead to a higher squalene yield, the potential of C5 redirection can be unleashed when combined with additional strategies to mitigate the negative impact on photosynthesis and endogenous terpene biosynthesis pathways. Such strategies include the integration of additional strategies to increase Calvin Benson cycle intermediates, the acceleration of Calvin Benson cycle intermediate regeneration, and the increased expression of DXPS for additional terpene intermediate biosynthesis. Moreover, the efficacy of C5 redirection could also be further enhanced with additional strategies to compartment the end products. One of such strategies is to use synthetic droplet to enhance terpene storage, as observed with previous work and evaluation to be effective, or increase fixation rate through addition of SBPase. Overall, the C5 redirection is effective, yet additional strategies are needed to mitigate its negative impact on photosynthesis and terpene biosynthesis.

V. CONCLUSIONS AND MOVING FORWARD

The initial goal to synthesize and compartmentalize squalene biosynthesis was successful to an extent when coupled to either ctpFPPS+ctpSQS or ctpFPPS+ctpSQS with the C2 photorespiratory bypass. Zhao Cheng's dissertation proves the double membrane of the chloroplast does indeed leak squalene, leading to increased phytosterols as well as reducing the squalene accumulation and subsequent extraction titer. Dr Zhao also designed a synthetic "oil droplet" to encapsulate squalene as it diffused out of the outer membrane. Using the C5 to C5 pathway to reroute sugars to squalene has seen a gain over the C2 photorespiratory bypass pathway while also less environmentally regulated. Regardless of the engineered carbon flux scheme and the intrinsic connections between that scheme and environmentally regulated endogenous carbon flux, carbon allocated to an engineered bioproduct must come at the expense of reduced carbon in the metabolism unless fixation efficiency is increased. More carbon fixed into the organism allows more carbon flux to be utilized in engineered pathways, and without an efficiency increase from genomic components, gain in photosynthetically fixed carbon will require a better fixation environment. Overall, conversion of largely abundant pentose sugars and pentose phosphates along with increased storage of squalene could provide more stable yields in tobacco. An ideal system would utilize the C5 design in a temporal, possibly induced state. When tobacco goes to flower it creates a strong sink to the emergent inflorescence, referred to by growers as "button" stage, greatly reducing carbon allocated to the desired terpene bioproduct. An inducible promoter system would allow full vegetative growth and allow the grower to decide the optimal time to harvest. Furthermore this approach would reduce the growth period to pre-flowering cultivation only, greatly reducing labor hours and time spent removing suckers. The application of this technique will require bioinformatics to determine a promoter with one of the following traits to

drive synthetic metabolism to terpenes: (1) A promoter that is activated by initiation of flowering or budding, or (2) A inducible promoter system that, when induced by an elicitor, is used to induce transcription of carbon flux and terpene sink after activation. One such system would be the AlcR/AlcA ethanol inducible promoter. This would eliminate growth phenotype differences between engineered and WT tobacco, then when plants and labor are ready, induced engineered scheme can function to rapidly convert sugars to terpene.

The application of biotechnology in metabolic engineering opens renewable and sustainable routes to bioproducts previously limited by natural biology. Application driven science often varies from theoretical and experimental sciences by introducing variables such as feasibility. The application of replacing shark liver oil derived squalene with a plant based source which does not compete with cooking oil markets for input, led to the production of engineered tobacco. This however is not the only bioproduct made in engineered tobacco, nor is it the only terpene these engineered carbon flux pathways were designed to produce. In the future, American farmers will be able plant tobacco as a cash crop once again. This tobacco will not go to consumers, but rather become the products we use every day.

CITATIONS

1. Paduch, R., et al., *Terpenes: substances useful in human healthcare.* 2007. **55**(5): p. 315.
2. Wang, X., D.R. Ort, and J.S.J.P.b.j. Yuan, *Photosynthetic terpene hydrocarbon production for fuels and chemicals.* 2015. **13**(2): p. 137-146.
3. Wang, X., et al., *Enhanced limonene production in cyanobacteria reveals photosynthesis limitations.* 2016. **113**(50): p. 14225-14230.
4. Ajikumar, P.K., et al., *Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli.* 2010. **330**(6000): p. 70-74.
5. Lichtenthaler, H.K.J.B.G.N.A.S., *Biosynthesis and accumulation of isoprenoid carotenoids and chlorophylls and emission of isoprene by leaf chloroplasts.* 2009. **3**: p. 81-94.
6. Rodríguez-Concepción, M.J.P.R., *Early steps in isoprenoid biosynthesis: multilevel regulation of the supply of common precursors in plant cells.* 2006. **5**(1): p. 1-15.
7. Cheng, A.X., et al., *Plant terpenoids: biosynthesis and ecological functions.* 2007. **49**(2): p. 179-186.
8. Zhao, L., et al., *Methylerythritol phosphate pathway of isoprenoid biosynthesis.* 2013. **82**: p. 497-530.
9. Handa, S., et al., *Mechanistic Studies of 1-Deoxy-D-Xylulose-5-Phosphate Synthase from Deinococcus radiodurans.* 2018. **4**(1).
10. Bartee, D. and C.L.J.A.o.c.r. Freel Meyers, *Toward Understanding the Chemistry and Biology of 1-Deoxy-d-xylulose 5-Phosphate (DXP) Synthase: A Unique Antimicrobial Target at the Heart of Bacterial Metabolism.* 2018. **51**(10): p. 2546-2555.

11. Murkin, A.S., K.A. Manning, and S.A.J.B.c. Kholodar, *Mechanism and inhibition of 1-deoxy-D-xylulose-5-phosphate reductoisomerase.* 2014. **57**: p. 171-185.

12. GLEIZES, M., et al., *Effects of light on terpene hydrocarbon synthesis in Pinus pinaster.* 1980. **50**(1): p. 16-20.

13. Tholl, D.J.C.o.i.p.b., *Terpene synthases and the regulation, diversity and biological roles of terpene metabolism.* 2006. **9**(3): p. 297-304.

14. Trapp, S.C. and R.B.J.G. Croteau, *Genomic organization of plant terpene synthases and molecular evolutionary implications.* 2001. **158**(2): p. 811-832.

15. Gimpel, J.A., et al., *Advances in microalgae engineering and synthetic biology applications for biofuel production.* 2013. **17**(3): p. 489-495.

16. Jiang, Z., et al., *Engineering triterpene and methylated triterpene production in plants provides biochemical and physiological insights into terpene metabolism.* 2016. **170**(2): p. 702-716.

17. Kempinski, C. and J.J.P.b.j. Chappell, *Engineering triterpene metabolism in the oilseed of Arabidopsis thaliana.* 2018.

18. Mahmoud, S.S. and R.B.J.P.o.t.N.A.o.S. Croteau, *Metabolic engineering of essential oil yield and composition in mint by altering expression of deoxyxylulose phosphate reductoisomerase and menthofuran synthase.* 2001. **98**(15): p. 8915-8920.

19. Martin, V.J., et al., *Engineering a mevalonate pathway in Escherichia coli for production of terpenoids.* 2003. **21**(7): p. 796.

20. Peralta-Yahya, P.P., et al., *Microbial engineering for the production of advanced biofuels.* 2012. **488**(7411): p. 320.

21. Fixen, K.R., et al., *Light-driven carbon dioxide reduction to methane by nitrogenase in a photosynthetic bacterium.* 2016. **113**(36): p. 10163-10167.

22. Shi, Y., et al., *Directed bioconversion of Kraft lignin to polyhydroxyalkanoate by Cupriavidus basilensis B-8 without any pretreatment.* 2017. **52**: p. 238-242.

23. Carlsson, A.S.J.B., *Plant oils as feedstock alternatives to petroleum–A short survey of potential oil crop platforms.* 2009. **91**(6): p. 665-670.

24. Spanova, M., G.J.E.j.o.l.s. Daum, and technology, *Squalene–biochemistry, molecular biology, process biotechnology, and applications.* 2011. **113**(11): p. 1299-1320.

25. Mcphee, D., et al., *Squalane from sugarcane.* 2014. **129**(6): p. 1-6.

26. Ölçer, H., J.C. Lloyd, and C.A.J.P.P. Raines, *Photosynthetic capacity is differentially affected by reductions in sedoheptulose-1, 7-bisphosphatase activity during leaf development in transgenic tobacco plants.* 2001. **125**(2): p. 982-989.

27. Miyagawa, Y., M. Tamoi, and S.J.N.b. Shigeoka, *Overexpression of a cyanobacterial fructose-1, 6-/sedoheptulose-1, 7-bisphosphatase in tobacco enhances photosynthesis and growth.* 2001. **19**(10): p. 965.

28. Lefebvre, S., et al., *Increased sedoheptulose-1, 7-bisphosphatase activity in transgenic tobacco plants stimulates photosynthesis and growth from an early stage in development.* 2005. **138**(1): p. 451-460.

29. Harrison, E.P., et al., *Reduced sedoheptulose-1, 7-bisphosphatase levels in transgenic tobacco lead to decreased photosynthetic capacity and altered carbohydrate accumulation.* 1997. **204**(1): p. 27-36.

30. Crafts-Brandner, S.J. and M.E.J.P.o.t.N.A.o.S. Salvucci, *Rubisco activase constrains the photosynthetic potential of leaves at high temperature and CO2.* 2000. **97**(24): p. 13430-13435.

31. Schrader, S.M., et al., *High temperature enhances inhibitor production but reduces fallover in tobacco Rubisco.* 2006. **33**(10): p. 921-929.

32. Muhaidat, R., R.F. Sage, and N.G.J.A.J.o.B. Dengler, *Diversity of Kranz anatomy and biochemistry in C4 eudicots.* 2007. **94**(3): p. 362-381.

33. Yamori, W., K. Hikosaka, and D.A.J.P.r. Way, *Temperature response of photosynthesis in C 3, C 4, and CAM plants: temperature acclimation and temperature adaptation.* 2014. **119**(1-2): p. 101-117.

34. Andrews, T. and H.J.P.P. Kane, Supplement, *Pyruvate is a by-product of Rubisco catalysis.* 1990. **93**(1): p. 9.

35. South, P.F., et al., *Synthetic glycolate metabolism pathways stimulate crop growth and productivity in the field.* 2019. **363**(6422): p. eaat9077.

36. Alexandrov, S.D.J.A.i.L.S., *Re-engineering of RuBisCO for the purpose of producing algal biofuels may be a lost cause.* 2016. **4**(1): p. 01-02.

37. Andrews, T.J., et al., *Rubisco: the consequences of altering its expression and activation in transgenic plants.* 1995: p. 1293-1300.

38. Morell, M.K., et al., *Rubisco: maladapted or misunderstood.* 1992. **40**(5): p. 431-441.

39. Huma, B., et al., *Stoichiometric analysis of the energetics and metabolic impact of photorespiration in C3 plants.* 2018.

40. Maurino, V.G. and C.J.C.o.i.p.b. Peterhansel, *Photorespiration: current status and approaches for metabolic engineering.* 2010. **13**(3): p. 248-255.

41. Kebeish, R., et al., *Chloroplastic photorespiratory bypass increases photosynthesis and biomass production in Arabidopsis thaliana.* 2007. **25**(5): p. 593.

42. Wright, L., et al., *1-Deoxyxylulose 5-phosphate synthase controls flux through the 2-C-methylerythritol 4-phosphate pathway in Arabidopsis thaliana.* 2014: p. pp. 114.245191.

43. Dudareva, N., et al., *Biosynthesis, function and metabolic engineering of plant volatile organic compounds.* 2013. **198**(1): p. 16-32.

44. Nagegowda, D.A.J.F.l., *Plant volatile terpenoid metabolism: biosynthetic genes, transcriptional regulation and subcellular compartmentation.* 2010. **584**(14): p. 2965-2973.

45. Pasoreck, E.K., et al., *Terpene metabolic engineering via nuclear or chloroplast genomes profoundly and globally impacts off-target pathways through metabolite signalling.* 2016. **14**(9): p. 1862-1875.

46. Zhao, C., et al., *Co-Compartmentation of Terpene Biosynthesis and Storage via Synthetic Droplet.* 2018. **7**(3): p. 774-781.

47. Banerjee, A. and T.J.N.p.r. Sharkey, *Methylerythritol 4-phosphate (MEP) pathway metabolic regulation.* 2014. **31**(8): p. 1043-1055.

48. Cordoba, E., M. Salmi, and P.J.J.o.E.B. León, *Unravelling the regulatory mechanisms that modulate the MEP pathway in higher plants.* 2009. **60**(10): p. 2933-2943.

49. Wu, S., et al., *Redirection of cytosolic or plastidic isoprenoid precursors elevates terpene production in plants.* 2006. **24**(11): p. 1441.

50. Wu, S., et al., *Engineering Triterpene Metabolism in Tobacco.* Planta, 2012. **236**(3): p. 867-877.

51. Wu, S.Q., et al., *Redirection of Cytosolic or Plastidic Isoprenoid Precursors Elevates Terpene Production in Plants.* Nature Biotechnology, 2006. **24**(11): p. 1441-1447.

52. Glen John Mcintyre, et al., *An Infinitely Expandable Cloning Strategy plus Repeat-Proof PCR for Working with Multiple shRNA.* PloS ONE, 2008. **3**(11): p. e3827. doi:10.1371/journal.pone.0003827.

53. R.B. Horsch, J.E.F., N.L. Hoffmann, D. Eichholtz, S.G. Rogers, R.T. Fraley, *A simple and general method for transferring genes into plants.* Science, 1985. **227**(4691): p. 1229-1231.

54. J. J. Doyle, J.L.D., *A rapid DNA isolation procedure for small quantities of fresh leaf tissue.* Phytochemical Bulletin, 1987. **19**(1): p. 11-15.

55. Yuan, J.S., et al., *Molecular and genomic basis of volatile-mediated indirect defense against insects in rice.* Plant Journal, 2008. **55**(3): p. 491-503.

56. Melvin J. Oliver, L.G., Danny C. Alexander, John A. Ryals, Bernard W.M. Wone, and a.J.C. Cushmand, *A Sister Group Contrast Using Untargeted Global Metabolomic Analysis Delineates the Biochemical Regulation Underlying Desiccation Tolerance in Sporobolus stapfianus.* Plant Cell, 2011. **23**(4): p. 1231-1248.

57. Ogren, W.L.J.A.R.o.P.P., *Photorespiration: pathways, regulation, and modification.* 1984. **35**(1): p. 415-442.

58. Powles, S.B., *Photoinhibition of Photosynthesis induced by visible light.* Annual Review of Plant Physiology, 1984(35): p. 27-28.

59. Cooper, M., I.J.T. DeLacy, and A. Genetics, *Relationships among analytical methods used to study genotypic variation and genotype-by-environment interaction in plant breeding multi-environment experiments.* 1994. **88**(5): p. 561-572.

60. Kozaki, A. and G.J.N. Takeba, *Photorespiration protects C3 plants from photooxidation.* 1996. **384**(6609): p. 557.

61. Betti, M., et al., *Manipulating photorespiration to increase plant productivity: recent advances and perspectives for crop improvement.* 2016. **67**(10): p. 2977-2988.

62. Sharkey, T.D.J.P.P., *Estimating the rate of photorespiration in leaves.* 1988. **73**(1): p. 147-152.

63. Pick, T.R., et al., *PLGG1, a plastidic glycolate glycerate transporter, is required for photorespiration and defines a unique class of metabolite transporters.* 2013. **110**(8): p. 3185-3190.

64. Kruger, N.J. and A.J.C.o.i.p.b. von Schaewen, *The oxidative pentose phosphate pathway: structure and organisation.* 2003. **6**(3): p. 236-246.

65. Melis, A.J.E. and E. Science, *Photosynthesis-to-fuels: from sunlight to hydrogen, isoprene, and botryococcene production.* 2012. **5**(2): p. 5531-5539.

66. Perez-Gil, J., et al., *Mutations in Escherichia coli aceE and ribB genes allow survival of strains defective in the first step of the isoprenoid biosynthesis pathway.* 2012. **7**(8): p. e43775.

67. Kirby, J., et al., *Host cells and methods for producing 1-deoxyxylulose 5-phosphate (DXP) and/or a DXP derived compound*. 2013, Google Patents.

68. Kirby, J., et al., *Enhancing terpene yield from sugars via novel routes to 1-deoxy-D-xylulose 5-phosphate.* 2015. **81**(1): p. 130-138.